

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 11:42:35 ; Search time 3139 Seconds
(without alignments)
482.111 Million cell updates/sec

Title: US-09-875-945-3

Perfect score: 52

Sequence: 1 ctcctgcattgccatattt.....gcagtaggtatctgtgcaca 52

Scoring table:

OMIGO-NUC-17

Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_em.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pi.*

25: em_ro.*

26: em_sts.*

27: em_un.*

28: em_vi.*

29: em_vt.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	100.0	70390	2	AC120044 Homo sapi
2	52	100.0	143409	9	AL162497 Human DNA
3	52	100.0	190740	2	AC011864 Homo sapi
4	20	38.5	203193	2	AC117652 Mus muscu
5	19	36.5	129361	9	AL672292 Human DNA
6	18	34.6	4513	10	AB060078 Mus muscu
7	18	34.6	73402	2	AC101535 Mus muscu
8	18	34.6	127029	9	AL445526 Human DNA
9	18	34.6	173249	2	AC123468 Rattus no
10	18	34.6	173804	2	AC122048 Mus muscu
11	18	34.6	174661	2	AC098609 Rattus no
12	18	34.6	175876	2	AC124465 Mus muscu
13	18	34.6	178183	2	AC095423 Rattus no
14	18	34.6	195574	10	AC091782 Genomic S
15	18	34.6	207558	2	AC114778 Homo sapi
16	18	34.6	209651	2	AC117596 Mus muscu
17	18	34.6	217132	2	AC107709 Mus muscu
18	18	34.6	223904	2	AL845373 Mus muscu
19	18	34.6	231542	2	AC120129 Mus muscu
20	17	32.7	68121	10	AP001917 Mus muscu
21	17	32.7	101340	2	AC100801 Homo sapi
22	17	32.7	103584	2	AC121702 Rattus no
23	17	32.7	108333	10	AP001293 Mus muscu
24	17	32.7	110000	2	AC098012 Continuation (3 of
25	17	32.7	110237	2	AF206725 Homo sapi
26	17	32.7	113114	9	AC079400 Homo sapi
27	17	32.7	132195	2	AC128824 Rattus no
28	17	32.7	139214	9	HS1128N12 Human DNA
29	17	32.7	139214	9	HS1128N12 Human DNA
30	17	32.7	144116	2	AC105959 Mus muscu
31	17	32.7	153177	2	AL353632 Homo sapi
32	17	32.7	156935	9	HSJ365019 Human DNA
33	17	32.7	161401	2	AC092087 Canis fam
34	17	32.7	174046	2	AC107517 Rattus no
35	17	32.7	197951	2	AC102731 Mus muscu
36	17	32.7	208548	2	AC092195 Canis fam
37	17	32.7	210423	2	AL672074 Mus muscu
38	17	32.7	211871	2	AC023248 Mus muscu
39	17	32.7	212265	2	AL844580 Mus muscu
40	17	32.7	214503	2	AL731842 Mus muscu
41	17	32.7	235197	2	AC122635 Rattus no
42	17	32.7	281000	10	MMU276505 AJ276505 Mus muscu
43	16	30.8	1636	8	SCYGF073C Z72858 S.cerevisia
44	16	30.8	1695	8	YSCSMD1A L04669 Saccharomyc
45	16	30.8	1708	8	NTNPP4 Z93771 N.tabacum n

ALIGNMENTS

RESULT 1
AC120044
LOCUS
DEFINITION Homo sapiens chromosome 15 clone RP11-163G3 map 15, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC120044
VERSION AC120044.2 GI:21327566
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 70390)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 15, clone RP11-163G3

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 70390)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Lakocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 70390)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 6, 2002 this sequence version replaced gi:20389713.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26755
Center clone name: 163_G_3

* NOTE: This record contains 87 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.

TITLE
JOURNAL
COMMENT

* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
675: contig of 675 bp in length
676 775: gap of 100 bp
776 1474: contig of 699 bp in length
1475 1574: gap of 100 bp
1575 2306: contig of 732 bp in length
2307 2406: gap of 100 bp
2407 3128: contig of 722 bp in length
3129 3228: gap of 100 bp
3229 3934: contig of 706 bp in length
3935 4034: gap of 100 bp
4035 4737: contig of 703 bp in length
4738 4837: gap of 100 bp
4838 5537: contig of 700 bp in length
5538 5637: gap of 100 bp
5638 6348: contig of 711 bp in length
6349 6448: gap of 100 bp
6449 7145: contig of 697 bp in length
7146 7245: gap of 100 bp
7246 7963: contig of 718 bp in length
7964 8063: gap of 100 bp
8064 8793: contig of 730 bp in length
8794 8893: gap of 100 bp
8894 9616: contig of 723 bp in length
9617 9716: gap of 100 bp
9717 10420: contig of 704 bp in length
10421 10520: gap of 100 bp
10521 11229: contig of 709 bp in length
11230 11329: gap of 100 bp
11330 12028: contig of 699 bp in length
12029 12128: gap of 100 bp
12129 12842: contig of 714 bp in length
12843 12942: gap of 100 bp
12943 13659: contig of 717 bp in length
13660 13759: gap of 100 bp
13760 14468: contig of 709 bp in length
14469 14568: gap of 100 bp
14569 15286: contig of 718 bp in length
15287 15386: gap of 100 bp
15387 16109: contig of 723 bp in length
16110 16209: gap of 100 bp
16210 16923: contig of 714 bp in length
16924 17023: gap of 100 bp
17024 17742: contig of 719 bp in length
17743 17842: gap of 100 bp
17843 18564: contig of 722 bp in length
18565 18664: gap of 100 bp
18665 19386: contig of 722 bp in length
19387 19486: gap of 100 bp
19487 20186: contig of 700 bp in length
20187 20286: gap of 100 bp
20287 20993: contig of 707 bp in length
20994 21093: gap of 100 bp
21094 21798: contig of 705 bp in length
21799 21898: gap of 100 bp
21899 22600: contig of 702 bp in length
22601 22700: gap of 100 bp
22701 23389: contig of 689 bp in length
23390 23489: gap of 100 bp
23490 24205: contig of 716 bp in length
24206 24305: gap of 100 bp
24306 25022: contig of 717 bp in length
25023 25122: gap of 100 bp
25123 25830: contig of 708 bp in length
25831 25930: gap of 100 bp
25931 26634: contig of 704 bp in length
26635 26734: gap of 100 bp
26735 27442: contig of 708 bp in length
27443 27542: gap of 100 bp
27543 28262: contig of 720 bp in length

```
* 28263 28362: gap of 100 bp
* 28363 29077: contig of 715 bp in length
* 29078 29177: gap of 100 bp
* 29178 29878: contig of 701 bp in length
* 29879 29978: gap of 100 bp
* 30682 30682: contig of 704 bp in length
* 30683 30782: gap of 100 bp
* 30783 31485: contig of 703 bp in length
* 31486 31585: gap of 100 bp
* 31586 32292: contig of 707 bp in length
* 32293 32392: gap of 100 bp
* 32393 33098: contig of 706 bp in length
* 33099 33198: gap of 100 bp
* 33199 33901: contig of 703 bp in length
* 33902 34001: gap of 100 bp
* 34002 34714: contig of 713 bp in length
* 34715 34814: gap of 100 bp
* 34815 35538: contig of 724 bp in length
* 35539 35638: gap of 100 bp
* 35639 36348: contig of 710 bp in length
* 36349 36448: gap of 100 bp
* 36449 37152: contig of 704 bp in length
* 37153 37252: gap of 100 bp
* 37253 37985: contig of 733 bp in length
* 37986 38085: gap of 100 bp
* 38086 38791: contig of 706 bp in length
* 38792 38891: gap of 100 bp
* 38892 39598: contig of 707 bp in length
* 39599 39698: gap of 100 bp
* 39699 40406: contig of 708 bp in length
* 40407 40506: gap of 100 bp
* 40507 41222: contig of 716 bp in length
* 41223 41322: gap of 100 bp
* 41323 42037: contig of 715 bp in length
* 42038 42137: gap of 100 bp
* 42138 42841: contig of 704 bp in length
* 42842 42941: gap of 100 bp
* 42942 43666: contig of 725 bp in length
* 43667 43766: gap of 100 bp
* 43767 44483: contig of 717 bp in length
* 44484 44583: gap of 100 bp

Query Match 100.08; Score 52; DB 2; Length 70390;
Best Local Similarity 100.08; Pred. No. 2.5e-22;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTGCATTGCGATTTGTGAGTCTACTTGCAGTAGTATCTGTGCACA 52
|||||
Db 68456 CTCCTGCATTGCGATTTGTGAGTCTACTTGCAGTAGTATCTGTGCACA 68507
```

```
RESULT 2
AL162497/c 143409 bp DNA linear PRI 06-JUN-2001
LOCUS Human DNA sequence from clone RP11-313L9 on chromosome 13, complete
DEFINITION sequence.
ACCESSION AL162497
VERSION AL162497.20 GI:14329908
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 143409)
Mashreghi-Mohammadi,M.
Direct Submission
Submitted (05-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Jun 8, 2001 this sequence version replaced gi:14280409.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
```

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; SW., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr13>
of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-313L9. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP11-313L9 is at 143409 in this sequence. The true left end of clone RP11-4066 is at 100074 in this sequence. The true right end of clone RP11-358F13 is at 100 in this sequence.

FEATURES

Location/Qualifiers	Source
1..143409	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="13"	
/clone="RP11-313L9"	
/clone_lib="RPCI-11.2"	
1..350	
/note="L1MA9 repeat: matches 5421. .5789 of consensus"	
351..654	
/note="L1PB1 repeat: matches 5846. .6155 of consensus"	
655..957	
/note="L1MA9 repeat: matches 5789. .6085 of consensus"	
2061..2106	
/note="23 copies 2 mer tg 91% conserved"	
2063..2106	
/note="11 copies 4 mer ttgtg 93% conserved"	
2138..2422	
/note="L1MD3 repeat: matches 6691. .7023 of consensus"	
2422..3051	
/note="L1MC4 repeat: matches 7168. .7837 of consensus"	
3674..4545	
/note="L1LP4 repeat: matches 5272. .6144 of consensus"	
5701..5910	
/note="7 copies 30 mer 89% conserved"	
6490..6613	
/note="MIR repeat: matches 92. .225 of consensus"	
9599..9648	
/note="25 copies 2 mer aa 94% conserved"	
10251..10559	
/note="AluX repeat: matches 2. .312 of consensus"	
11134..11521	
/note="L1MB7 repeat: matches 5762. .6161 of consensus"	
12667..12710	
/note="11 copies 4 mer ttgtg 81% conserved"	
13210..13309	
/note="L2 repeat: matches 2559. .2674 of consensus"	
13794..14148	
/note="THE1C repeat: matches 1. .371 of consensus"	
14816..15030	
/note="MIR repeat: matches 49. .256 of consensus"	
14982..15037	
/note="L2 repeat: matches 2651. .2705 of consensus"	

```

repeat_region 15476..15940
/Note="L2 repeat: matches 1314..1835 of consensus"
repeat_region 16305..16609
/Note="AlusX repeat: matches 1..300 of consensus"
repeat_region 16683..16722
/Note="20 copies 2 mer tt 77% conserved"
repeat_region 16948..17073
/Note="MLRIC repeat: matches 348..464 of consensus"
repeat_region 17074..17360
/Note="AluJo repeat: matches 13..299 of consensus"
repeat_region 17361..17661
/Note="MLRIC repeat: matches 7..348 of consensus"
repeat_region 17974..18347
/Note="L1MD1 repeat: matches 5659..6029 of consensus"
repeat_region 18399..18554
/Note="3 copies 52 mer 75% conserved"
repeat_region 18400..18549
/Note="5 copies 30 mer 73% conserved"
repeat_region 18413..18552
/Note="5 copies 28 mer 74% conserved"
repeat_region 18565..18738
/Note="L1MD2 repeat: matches 6150..6335 of consensus"
repeat_region 19070..19105
/Note="9 copies 4 mer gata 88% conserved"
repeat_region 19363..19664
/Note="L2 repeat: matches 2436..2750 of consensus"
repeat_region 20032..20053
/Note="11 copies 2 mer aa 100% conserved"
repeat_region 21272..21554
/Note="Alusq repeat: matches 1..283 of consensus"
repeat_region 22754..22918
/Note="3 copies 55 mer 77% conserved"
repeat_region 26601..26684
/Note="L2 repeat: matches 2668..2748 of consensus"
repeat_region 26685..26814
/Note="AlubB repeat: matches 2..131 of consensus"
repeat_region 26815..27110
/Note="Alusq repeat: matches 1..303 of consensus"
repeat_region 27111..27285
/Note="AlubB repeat: matches 130..307 of consensus"
repeat_region 27286..27578
/Note="L2 repeat: matches 2359..2668 of consensus"
repeat_region 29945..30257
/Note="AluY repeat: matches 1..306 of consensus"
repeat_region 30812..31078
/Note="AlusX repeat: matches 1..267 of consensus"
repeat_region 31887..32091
/Note="WERS8A repeat: matches 1..222 of consensus"
repeat_region 32818..32853
/Note="9 copies 4 mer acac 100% conserved"
repeat_region 32856..32957
/Note="WERS7C repeat: matches 185..289 of consensus"
repeat_region 32958..33406
/Note="MLRIC repeat: matches 1..466 of consensus"
repeat_region 33407..33867
/Note="WERS7C repeat: matches 289..726 of consensus"
repeat_region 33895..34718
/Note="8 copies 103 mer 61% conserved"
repeat_region 33915..34519
/Note="121 copies 5 mer atata 56% conserved"
repeat_region 33920..34103
/Note="92 copies 2 mer at 57% conserved"
repeat_region 33924..34553
/Note="21 copies 30 mer 61% conserved"
repeat_region 33925..34730
/Note="13 copies 62 mer 61% conserved"
repeat_region 33940..34709
/Note="14 copies 55 mer 61% conserved"
repeat_region 33954..34213
/Note="13 copies 20 mer 70% conserved"
repeat_region 33999..34698
/Note="25 copies 28 mer 58% conserved"
repeat_region 34014..34650

```

```

/Note="13 copies 49 mer 62% conserved"
34071..34642
/Note="11 copies 52 mer 62% conserved"
34300..34709
/Note="205 copies 2 mer aa 56% conserved"
34390..34689
/Note="15 copies 20 mer 60% conserved"
34488..34699
/Note="53 copies 4 mer taaa 59% conserved"
34512..34703
/Note="64 copies 3 mer ata 60% conserved"
34629..34718
/Note="3 copies 30 mer 78% conserved"
35062..35647
/Note="L1ME1 repeat: matches 5550..6144 of consensus"
35694..35885
/Note="L2 repeat: matches 2054..2250 of consensus"
35882..36229
/Note="L2 repeat: matches 1446..1832 of consensus"
36377..36484
/Note="MIR repeat: matches 48..147 of consensus"
39838..39877
/Note="20 copies 2 mer tt 82% conserved"
39841..39875
/Note="7 copies 5 mer tgttt 85% conserved"
41466..41514
/Note="L2 repeat: matches 2699..2747 of consensus"
41792..42143
/Note="THE1B repeat: matches 1..364 of consensus"
43471..43590
/Note="L2 repeat: matches 2629..2750 of consensus"
44937..45157
/Note="L2 repeat: matches 2259..2489 of consensus"
45491..45781

```

Query Match 100.0%; Score 52; DB 9; Length 143409;
 Best Local Similarity 100.0%; Pred. No. 2.5e-22;

Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCGCATTCGCATATTTGTGAGTGCACCTGCAGTAGGTAGTATCTGTCACA 52

Db 135485 CTCCTCGCATTCGCATATTTGTGAGTGCACCTGCAGTAGGTAGTATCTGTCACA 135434

RESULT 3

AC011864

LOCUS AC011864 190740 bp DNA linear HTG 12-MAR-2000

DEFINITION Homo sapiens clone RP11-16C4, WORKING DRAFT SEQUENCE, 51 unordered pieces.

ACCESSION AC011864

VERSION AC011864.3 GI:7230118

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 190740)

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Homo sapiens, clone RP11-16C4

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 190740)

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baldwin, J., Barna, N., Becker, R., Boguslavsky, L., Boukhalter, B.,

Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,

Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,

Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,

Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Lehoczky, J., Liu, C., Locke, K., Macdonald, P., Marquis, N.,

McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,

Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,

Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-OCT-1992) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6479010.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3517
Center clone name: 16_C_4

----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 121051 bases at least Q40
Consensus quality: 152215 bases at least Q30
Consensus quality: 171265 bases at least Q20
Insert size: 167000; agarose-fp
Insert size: 185740; sum-of-contigs
Quality coverage: 3.5 in Q20 bases; agarose-fp
Quality coverage: 3.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 51 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
* 1198 1297: gap of 100 bp in length
* 1298 2601: contig of 1304 bp in length
* 2602 2701: gap of 100 bp
* 2702 3906: contig of 1205 bp in length
* 3907 4006: gap of 100 bp
* 4007 5023: contig of 1017 bp in length
* 5024 5123: gap of 100 bp
* 5124 6559: contig of 1436 bp in length
* 6560 6659: gap of 100 bp
* 6660 8178: contig of 1519 bp in length
* 8179 8278: gap of 100 bp
* 8279 9354: contig of 1076 bp in length
* 9355 9454: gap of 100 bp
* 9455 10758: contig of 1304 bp in length
* 10759 10858: gap of 100 bp
* 10859 12011: contig of 1153 bp in length
* 12012 12111: gap of 100 bp
* 12112 13508: contig of 1397 bp in length
* 13509 13608: gap of 100 bp
* 13609 14887: contig of 1279 bp in length
* 14888 14987: gap of 100 bp
* 14988 16156: contig of 1169 bp in length
* 16157 16256: gap of 100 bp
* 16257 17785: contig of 1529 bp in length
* 17786 17885: gap of 100 bp
* 17886 18947: contig of 1062 bp in length
* 18948 19047: gap of 100 bp
* 19048 20330: contig of 1283 bp in length
* 20331 20430: gap of 100 bp
* 20431 21914: contig of 1484 bp in length
* 21915 22014: gap of 100 bp
* 22015 23101: contig of 1087 bp in length
* 23102 23201: gap of 100 bp
* 23202 24447: contig of 1246 bp in length
* 24448 24547: gap of 100 bp

* 24548 26032: contig of 1485 bp in length
* 26033 26132: gap of 100 bp
* 26133 28012: contig of 1880 bp in length
* 28013 28112: gap of 100 bp
* 28113 30280: contig of 2168 bp in length
* 30281 30380: gap of 100 bp
* 30381 31436: contig of 1056 bp in length
* 31437 31536: gap of 100 bp
* 31537 33906: contig of 2370 bp in length
* 33907 34006: gap of 100 bp
* 34007 35444: contig of 1438 bp in length
* 35445 35544: gap of 100 bp
* 35545 38284: contig of 2740 bp in length
* 38285 38384: gap of 100 bp
* 38385 40159: contig of 1775 bp in length
* 40160 40259: gap of 100 bp
* 40260 42089: contig of 1830 bp in length
* 42090 42189: gap of 100 bp
* 42190 43894: contig of 1705 bp in length
* 43895 43994: gap of 100 bp
* 43995 46499: contig of 2505 bp in length
* 46500 46599: gap of 100 bp
* 46600 48033: contig of 1434 bp in length
* 48034 48133: gap of 100 bp
* 48134 51777: contig of 3644 bp in length
* 51778 51877: gap of 100 bp
* 51878 55116: contig of 3239 bp in length
* 55117 55216: gap of 100 bp
* 55217 57721: contig of 2505 bp in length
* 57722 57821: gap of 100 bp
* 57822 61349: contig of 3528 bp in length
* 61350 61449: gap of 100 bp
* 61450 65218: contig of 3769 bp in length
* 65219 65318: gap of 100 bp
* 65319 68743: contig of 3425 bp in length
* 68744 68843: gap of 100 bp
* 68844 74044: contig of 5201 bp in length
* 74045 74144: gap of 100 bp
* 74145 78742: contig of 4598 bp in length
* 78743 78842: gap of 100 bp
* 78843 82807: contig of 3965 bp in length
* 82808 82907: gap of 100 bp
* 82908 87683: contig of 4776 bp in length
* 87684 87783: gap of 100 bp
* 87784 92706: contig of 4923 bp in length
* 92707 92806: gap of 100 bp
* 92807 99946: contig of 7140 bp in length
* 99947 100046: gap of 100 bp
* 100047 107272: contig of 7226 bp in length
* 107273 107372: gap of 100 bp
* 107373 115832: contig of 8460 bp in length
* 115833 115932: gap of 100 bp
* 115933 124872: contig of 8940 bp in length
* 124873 124972: gap of 100 bp
* 124973 134820: contig of 9848 bp in length
* 134821 134920: gap of 100 bp
* 134921 143495: contig of 8575 bp in length
* 143496 143595: gap of 100 bp
* 143596 152927: contig of 9332 bp in length
* 152928 153027: gap of 100 bp
* 153028 164076: contig of 11049 bp in length
* 164077 164176: gap of 100 bp
* 164177 173869: contig of 9693 bp in length
* 173870 173969: gap of 100 bp
* 173970 190740: contig of 16771 bp in length.

FEATURES
source

1. .190740
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Rp11-16C4"
/clone_lib="RPCI-11 Human Male BAC"
1. .1197
/note="assembly_fragment"

misc_feature

```

misc_feature      1298..2601
                  /note="assembly_fragment"
misc_feature      2702..3906
                  /note="assembly_fragment"
misc_feature      4007..5023
                  /note="assembly_fragment"
misc_feature      5124..6559
                  /note="assembly_fragment"
misc_feature      6660..8178
                  /note="assembly_fragment"
misc_feature      8279..9354
                  /note="assembly_fragment"
misc_feature      9455..10758
                  /note="assembly_fragment"
misc_feature      10859..12011
                  /note="assembly_fragment"
misc_feature      12112..13508
                  /note="assembly_fragment"
misc_feature      13609..14887
                  /note="assembly_fragment"
misc_feature      14988..16156
                  /note="assembly_fragment"
misc_feature      16257..17785
                  /note="assembly_fragment"
misc_feature      17886..18947
                  /note="assembly_fragment"
misc_feature      19048..20330
                  /note="assembly_fragment"

```

```

Query Match      100.0%; Score 52; DB 2; Length 190740;
Best Local Similarity 100.0%; Pred. No. 2.5e-22;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 CTTCCGCGATTCGCCATATTTGTGAGGTGACCTTGCGAGTAGGTCGTGCACA 52
    |||||
Db 68067 CTTCCGCGATTCGCCATATTTGTGAGGTGACCTTGCGAGTAGGTCGTGCACA 68118

```

```

RESULT 4
AC117652/c
LOCUS              AC117652              203193 bp      DNA      linear      HTG 06-AUG-2002
DEFINITION        Mus musculus clone RP23-291D3, WORKING DRAFT SEQUENCE, 41 ordered
                    pieces.
ACCESSION          AC117652
VERSION            AC117652.2  GI:22123330
KEYWORDS            HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE              house mouse.
ORGANISM            Mus musculus
                    Eukaryota; Metazoa; Chordata; Craniata; vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE           Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,A., Allen,N.,
AUTHORS            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
                    Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
                    Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
                    Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
                    Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
                    Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
                    Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
                    Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
                    Landers,T., Lehoczký,J., Levine,R., Lindblad-Toh,K., Liu,G.,
                    MacLean,C., Macdonald,P., Major,J., Marquis,N., Meneus,L.,
                    McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
                    Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
                    Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
                    Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
                    Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
                    Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
                    Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
                    Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

```

```

Topham,K., Travers,M., Travis,N., Travis,N., Trigillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,A. and Zody,M.
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 203193)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,A., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 6, 2002 this sequence version replaced gi:20128416.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23684
Center clone name: 291_D_3
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 190811 bases at least Q40
Consensus quality: 196133 bases at least Q30
Consensus quality: 197994 bases at least Q20
Insert size: 230000; agarose-fp
Insert size: 199193; sum-of-contigs
Quality coverage: 4.1 in Q20 bases; agarose-fp
Quality coverage: 4.8 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 41 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 956: contig of 956 bp in length
* 957 1056: gap of 100 bp
* 1057 1812: contig of 756 bp in length
* 1813 1912: gap of 100 bp
* 1913 2571: contig of 659 bp in length
* 2572 2671: gap of 100 bp
* 2672 3534: contig of 863 bp in length
* 3535 3634: gap of 100 bp
* 3635 4507: contig of 873 bp in length
* 4508 4607: gap of 100 bp
* 4608 5874: contig of 1267 bp in length
* 5875 5974: gap of 100 bp

```

TITLE JOURNAL REFERENCE AUTHORS

TITLE JOURNAL COMMENT

* 5975 7020: contig of 1046 bp in length
* 7021 7120: gap of 100 bp
* 7121 8380: contig of 1260 bp in length
* 8381 8480: gap of 100 bp
* 8481 9616: contig of 1136 bp in length
* 9617 9716: gap of 100 bp
* 9717 10980: contig of 1264 bp in length
* 10981 11080: gap of 100 bp
* 11081 12896: contig of 1816 bp in length
* 12897 12996: gap of 100 bp
* 12997 15433: contig of 2437 bp in length
* 15434 15533: gap of 100 bp
* 15534 17158: contig of 1625 bp in length
* 17159 17258: gap of 100 bp
* 17259 19061: contig of 1803 bp in length
* 19062 19161: gap of 100 bp
* 19162 21017: contig of 1856 bp in length
* 21018 21117: gap of 100 bp
* 21118 22665: contig of 1548 bp in length
* 22666 22765: gap of 100 bp
* 22766 24693: contig of 1928 bp in length
* 24694 24793: gap of 100 bp
* 24794 27131: contig of 2338 bp in length
* 27132 27231: gap of 100 bp
* 27232 29568: contig of 2337 bp in length
* 29569 29668: gap of 100 bp
* 29669 31302: contig of 1634 bp in length
* 31303 31402: gap of 100 bp
* 31403 35043: contig of 3641 bp in length
* 35044 35143: gap of 100 bp
* 35144 38234: contig of 3091 bp in length
* 38235 38334: gap of 100 bp
* 38335 42932: contig of 4598 bp in length
* 42933 43032: gap of 100 bp
* 43033 46231: contig of 3199 bp in length
* 46232 46331: gap of 100 bp
* 46332 50239: contig of 3908 bp in length
* 50240 50339: gap of 100 bp
* 50340 54474: contig of 4135 bp in length
* 54475 54574: gap of 100 bp
* 54575 59666: contig of 5092 bp in length
* 59667 59766: gap of 100 bp
* 59767 64692: contig of 4926 bp in length
* 64693 64792: gap of 100 bp
* 64793 72399: contig of 7607 bp in length
* 72400 72499: gap of 100 bp
* 72500 79793: contig of 7294 bp in length
* 79794 79893: gap of 100 bp
* 79894 86579: contig of 6666 bp in length
* 86580 86679: gap of 100 bp
* 86680 93045: contig of 6366 bp in length
* 93046 93145: gap of 100 bp
* 93146 107126: contig of 13981 bp in length
* 107127 107226: gap of 100 bp
* 107227 119453: contig of 12227 bp in length
* 119454 119553: gap of 100 bp
* 119554 130493: contig of 10940 bp in length
* 130494 130593: gap of 100 bp
* 130594 139561: contig of 9368 bp in length
* 139562 140061: gap of 100 bp
* 140062 154794: contig of 14733 bp in length
* 154795 154894: gap of 100 bp
* 154895 168254: contig of 13360 bp in length
* 168255 168354: gap of 100 bp
* 168355 180509: contig of 12155 bp in length
* 180510 180609: gap of 100 bp
* 180610 195305: contig of 14696 bp in length
* 195306 195405: gap of 100 bp
* 195406 203193: contig of 7788 bp in length.

FEATURES
Source

1. 203193
/organism="Mus musculus"
/db_xref="taxon:10090"

FEATURES

Location/Qualifiers

misc_feature /clone="RP23-291D3"
1. .956
/clone_lib="RPC1-23 Female Mouse BAC"
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
misc_feature 1057..1812
/note="assembly_fragment"
misc_feature 1913..2571
/note="assembly_fragment"
misc_feature 2672..3534
/note="assembly_fragment"
misc_feature 3635..4507
/note="assembly_fragment"
misc_feature 4608..5874
/note="assembly_fragment"
misc_feature 5975..7020
/note="assembly_fragment"
misc_feature 7121..8380
/note="assembly_fragment"

Query Match 38.5%; Score 20; DB 2; Length 203193;
Best Local Similarity 100.0%; Pred.No.0.096;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 CCATATTTGTGAGGTCACCT 32
|||||
Db 111997 CCATATTTGTGAGGTCACCT 111978

RESULT 5
AL672292/c

LOCUS AL672292 129361 bp DNA linear PRI 22-MAY-2002
DEFINITION Human DNA sequence from clone Xxyac-39EC11 on chromosome 6,
complete sequence.

ACCESSION AL672292
VERSION AL672292.9 GI:21212341

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 129361)
Direct Submission

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (21-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
On May 25, 2002 this sequence version replaced gi:20338546.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
XXYac-39EC11 is from the ICI Human YAC library (RA) VECTOR: pYAC4.
Location/Qualifiers

```

source
1. .129361
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="XXyac-39C11"
/clone_lib="ICI_YAC_RA"
BASE COUNT 36825 a 25923 c 26315 g 40298 t
ORIGIN

Query Match 36.5%; Score 19; DB 9; Length 129361;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 TTGTGAGGTCACTGCAG 36
|||||
Db 27647 TTGTGAGGTCACTGCAG 27629

RESULT 6
AB060078/c 4513 bp DNA linear ROD 16-OCT-2001
LOCUS AB060078 Mus musculus gene for ghrelin, complete cds.
DEFINITION AB060078
ACCESSION AB060078
VERSION AB060078.1 GI:16151744
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Tanaka,M., Hayashida,Y., Iguchi,T., Nakao,N., Nakai,N. and
Nakashima,K.
TITLE
Organization of mouse ghrelin gene
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 4513)
TANAKA,M.
AUTHORS
Direct Submission
TITLE
Submitted (13-APR-2001) Minoru Tanaka, Mie university Faculty of
Medicine, department of Biochemistry; 2-174 Edobashi, Tsu, Mie
514-8507, Japan (E-mail:mi-tanaka@doc.medic.mie-u.ac.jp,
Tel:+81-59-231-5007(ex.6347), Fax:+81-59-231-5007)
JOURNAL
Location/Qualifiers
FEATURES
source
1. .4513
/organism="Mus musculus"
/db_xref="taxon:10090"
736..742
TATA_signal
CDS
join(1219..1326,1420..1536,3479..3587,4371..4390)
/codon_start=1
/product="ghrelin"
/protein_id="BAB69857.1"
/db_xref="GI:16151745"
/translation="MLSSGTICSLLLSLMLMDMAMAGSSFLSPFHOKAQRKESKPP
PAKIQRALEGNLHPDRGQAEETEELEIFRNAPFDVGILKSAQYQQHGRALGKPL
QDIWEEVKERAPAK"
BASE COUNT 1315 a 1212 c 1007 g 979 t
ORIGIN

Query Match 34.6%; Score 18; DB 10; Length 4513;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 ATTGCCATATTTGTGAGG 26
|||||
Db 4150 ATTGCCATATTTGTGAGG 4133

RESULT 7
AC101535
LOCUS AC101535 73402 bp DNA linear HTG 23-NOV-2001
DEFINITION AC101535 Mus musculus clone RP23-192D22, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC101535
VERSION AC101535.1 GI:17060310
KEYWORDS HTG; HTGS_PHASE0.

Mus musculus.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 73402)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-192D22
Unpublished
2 (bases 1 to 73402)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukngalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferrelira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kellis,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WTBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L16914
Center clone name: 192_D_22
-----
* NOTE: This record contains 91 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 680: contig of 680 bp in length
* 681 780: gap of 100 bp
* 781 1502: contig of 722 bp in length
* 1503 1602: gap of 100 bp
* 1603 2318: contig of 716 bp in length
* 2319 2418: gap of 100 bp
* 2419 3136: contig of 718 bp in length
* 3137 3236: gap of 100 bp
* 3237 3929: contig of 693 bp in length
* 3930 4029: gap of 100 bp
* 4030 4718: contig of 689 bp in length
* 4719 4818: gap of 100 bp
* 4819 5498: contig of 680 bp in length
* 5499 5598: gap of 100 bp
* 5599 6311: contig of 713 bp in length
* 6312 6411: gap of 100 bp

```

```
* 6412 7129: contig of 718 bp in length
* 7130 7229: gap of 100 bp
* 7230 7939: contig of 710 bp in length
* 7940 8039: gap of 100 bp
* 8040 8736: contig of 697 bp in length
* 8737 8836: gap of 100 bp
* 8837 9530: contig of 694 bp in length
* 9531 9630: gap of 100 bp
* 9631 10320: contig of 690 bp in length
* 10321 10420: gap of 100 bp
* 10421 11147: contig of 727 bp in length
* 11148 11247: gap of 100 bp
* 11248 11957: contig of 710 bp in length
* 11958 12057: gap of 100 bp
* 12058 12756: contig of 699 bp in length
* 12757 12856: gap of 100 bp
* 12857 13553: contig of 697 bp in length
* 13554 13653: gap of 100 bp
* 13654 14347: contig of 694 bp in length
* 14348 14447: gap of 100 bp
* 14448 15171: contig of 724 bp in length
* 15172 15271: gap of 100 bp
* 15272 15981: contig of 710 bp in length
* 15982 16081: gap of 100 bp
* 16082 16766: contig of 685 bp in length
* 16767 16866: gap of 100 bp
* 16867 17568: contig of 702 bp in length
* 17569 17668: gap of 100 bp
* 17669 18361: contig of 693 bp in length
* 18362 18461: gap of 100 bp
* 18462 19130: contig of 669 bp in length
* 19131 19230: gap of 100 bp
* 19231 19955: contig of 725 bp in length
* 19956 20055: gap of 100 bp
* 20056 20774: contig of 719 bp in length
* 20775 20874: gap of 100 bp
* 20875 21592: contig of 718 bp in length
* 21593 21692: gap of 100 bp
* 21693 22408: contig of 716 bp in length
* 22409 22508: gap of 100 bp
* 22509 23212: contig of 704 bp in length
* 23213 23312: gap of 100 bp
* 23313 24015: contig of 703 bp in length
* 24016 24115: gap of 100 bp
* 24116 24836: contig of 721 bp in length
* 24837 24936: gap of 100 bp
* 24937 25665: contig of 729 bp in length
* 25666 25765: gap of 100 bp
* 25766 26466: contig of 701 bp in length
* 26467 26566: gap of 100 bp
* 26567 27271: contig of 705 bp in length
* 27272 27371: gap of 100 bp
* 27372 28094: contig of 723 bp in length
* 28095 28194: gap of 100 bp
* 28195 28889: contig of 695 bp in length
* 28890 28989: gap of 100 bp
* 28990 29717: contig of 728 bp in length
* 29718 29817: gap of 100 bp
* 29818 30545: contig of 728 bp in length
* 30546 30645: gap of 100 bp
* 30646 31340: contig of 695 bp in length
* 31341 31440: gap of 100 bp
* 31441 32132: contig of 692 bp in length
* 32133 32232: gap of 100 bp
* 32233 32945: contig of 713 bp in length
* 32946 33045: gap of 100 bp
* 33046 33742: contig of 697 bp in length
* 33743 33842: gap of 100 bp
* 33843 34557: contig of 715 bp in length
* 34558 34657: gap of 100 bp
* 34658 35364: contig of 707 bp in length
* 35365 35464: gap of 100 bp
* 35465 36165: contig of 701 bp in length
```

```
* 36166 36265: gap of 100 bp
* 36266 37000: contig of 735 bp in length
* 37001 37100: gap of 100 bp
* 37101 37797: contig of 697 bp in length
* 37798 37897: gap of 100 bp
* 37898 38632: contig of 735 bp in length
* 38633 38732: gap of 100 bp
* 38733 39436: contig of 704 bp in length
* 39437 39536: gap of 100 bp
* 39537 40259: contig of 723 bp in length
* 40260 40359: gap of 100 bp
* 40360 41077: contig of 718 bp in length
* 41078 41177: gap of 100 bp
* 41178 41876: contig of 699 bp in length
* 41877 41976: gap of 100 bp
* 41977 42689: contig of 713 bp in length
* 42690 42789: gap of 100 bp
* 42790 43478: contig of 689 bp in length
* 43479 43578: gap of 100 bp
* 43579 44279: contig of 701 bp in length
* 44280 44379: gap of 100 bp
* 44380 45084: contig of 705 bp in length
* 45085 45184: gap of 100 bp
* 45185 45882: contig of 698 bp in length
* 45883 45982: gap of 100 bp
* 45983 46708: contig of 726 bp in length
* 46709 46808: gap of 100 bp
* 46809 47507: contig of 699 bp in length
* 47508 47607: gap of 100 bp
* 47608 48323: contig of 716 bp in length
* 48324 48423: gap of 100 bp
* 48424 49097: contig of 674 bp in length
* 49098 49197: gap of 100 bp
* 49198 49892: contig of 695 bp in length
* 49893 49992: gap of 100 bp
* 49993 50703: contig of 711 bp in length
* 50704 50803: gap of 100 bp
* 50804 51507: contig of 704 bp in length
* 51508 51607: gap of 100 bp
* 51608 52318: contig of 711 bp in length
* 52319 52418: gap of 100 bp
* 52419 53139: contig of 721 bp in length
* 53140 53239: gap of 100 bp
* 53240 53971: contig of 732 bp in length
* 53972 54071: gap of 100 bp
* 54072 54756: contig of 685 bp in length
* 54757 54856: gap of 100 bp
* 54857 55575: contig of 719 bp in length
* 55576 55675: gap of 100 bp
```

Query Match 34.6%; Score 18; DB 2; Length 73402;
Best Local Similarity 100.0%; Pred. No. 1.9;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAGGTCACCTGCAGTAGG 40

|||||

Db 51785 GAGGTCACCTGCAGTAGG 51802

RESULT 8

AL445526

LOCUS

DEFINITION

Human DNA sequence from clone RP11-3212 on chromosome 9 Contains

STSs and GSSs, complete sequence.

ACCESSION

AL445526

VERSION

AL445526.9

GI:12331131

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 127029)

AUTHORS

Hammond,S.

linear PRI 26-SEP-2001

TITLE Direct Submission
JOURNAL Submitted (24-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT requests: clonerequest@sanger.ac.uk
 On Jan 22, 2001 this sequence version replaced gi:12329469. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em', EMBL; Sw', SWISSPROT; Tr', TREMBL; Wp', WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>
 Rp11-3212 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
FEATURES **source** 1. 127029
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /clone="RP11-3212"
 /clone.lib="RPCI-11.1"
 /note="match: GSS: Em:AQ532742"
 complement(39..568)
 /note="match: GSS: Em:AQ619914"
 complement(160..557)
 /note="match: GSS: Em:AQ404394"
 1635..1742
 /note="27 copies 4 mer aaag 63% conserved"
 1646..2020
 /note="match: STS: Em:HSA050YH1"
 1804..1859
 /note="28 copies 2 mer ca 91% conserved"
 1807..1858
 /note="13 copies 4 mer acac 94% conserved"
 complement(2528..2974)
 /note="match: GSS: Em:AQ461208"
 6076..6357
 /note="match: GSS: Em:AQ745065"
 6078..6189
 /note="56 copies 2 mer ca 83% conserved"
 6081..6208
 /note="32 copies 4 mer acac 80% conserved"
 6894..6943
 /note="25 copies 2 mer gt 84% conserved"
 6896..6943
 /note="12 copies 4 mer gtgt 85% conserved"
 complement(8258..8736)
 /note="match: GSS: Em:AQ471124"

misc_feature complement(8286..8754)
 repeat_region /note="match: GSS: Em:AQ471657"
 repeat_region 9868..10300
 /note="TIGGER2 repeat: matches 2273..2718 of consensus"
 10291..10357
 /note="TIGGER2 repeat: matches 1..69 of consensus"
 repeat_region 10358..10393
 /note="9 copies 4 mer tgtg 86% conserved"
 repeat_region 10412..10859
 /note="MLT2B repeat: matches 1..442 of consensus"
 repeat_region 10860..12212
 /note="HERVL repeat: matches 3064..4434 of consensus"
 repeat_region 12228..12284
 /note="MLT2B repeat: matches 394..448 of consensus"
 repeat_region 12285..12376
 /note="23 copies 4 mer tata 79% conserved"
 repeat_region 12382..12768
 /note="MLT2B repeat: matches 1..388 of consensus"
 repeat_region 12769..12808
 /note="20 copies 2 mer ca 100% conserved"
 repeat_region 12905..12932
 /note="7 copies 4 mer caat 92% conserved"
 misc_feature complement(12986..13298)
 repeat_region /note="match: GSS: Em:AQ244408"
 13344..14502
 /note="TIGGER1 repeat: matches 26..997 of consensus"
 repeat_region 14518..14887
 /note="TIGGER1 repeat: matches 1103..1465 of consensus"
 repeat_region 15189..15223
 /note="TIGGER1 repeat: matches 1465..1499 of consensus"
 repeat_region 15222..15889
 /note="TIGGER1 repeat: matches 1677..2359 of consensus"
 misc_feature complement(17899..18296)
 repeat_region /note="match: GSS: Em:AQ462100"
 18086..18577
 /note="WER34 repeat: matches 17..543 of consensus"
 misc_feature 18317..18823
 /note="match: GSS: Em:AQ413329"
 21864..22347
 /note="match: GSS: Em:AQ245163"
 21865..22481
 /note="match: GSS: Em:B57007"
 22567..22804
 /note="119 copies 2 mer aa 55% conserved"
 misc_feature complement(25607..26291)
 misc_feature /note="match: GSS: Em:AQ488229"
 complement(25778..26286)
 /note="match: GSS: Em:AQ420662"
 28469..28544
 /note="19 copies 4 mer atat 71% conserved"
 repeat_region 28470..28547
 /note="39 copies 2 mer ta 70% conserved"
 repeat_region 30718..30749
 /note="8 copies 4 mer atat 87% conserved"
 repeat_region 32647..33174
 /note="MLT2A repeat: matches 1..453 of consensus"
 misc_feature complement(33782..34303)
 misc_feature /note="match: GSS: Em:AQ431958"
 complement(34211..34702)
 /note="match: GSS: Em:AQ077619"
 complement(34235..34696)
 /note="match: GSS: Em:AQ877892"
 34314..34749
 /note="match: GSS: Em:AQ880845"
 34416..34845
 /note="match: GSS: Em:AQ633594"
 repeat_region 35386..35429
 /note="11 copies 4 mer caca 81% conserved"
 repeat_region 37899..37948
 /note="25 copies 2 mer aa 72% conserved"
 repeat_region 38569..38644
 /note="38 copies 2 mer aa 65% conserved"
 repeat_region 39925..39968

```

/misc_feature      /note="22 copies 2 mer ct 93% conserved"
44526. .45251
/note="match: GSS: Em:AQ896493"
/misc_feature      44558. .45028
/note="match: GSS: Em:AQ694275"
45716. .45804
/note="LIM48 repeat: matches 5888. .5970 of consensus"
45890. .46448
/note="MER39 repeat: matches 13. .548 of consensus"
46485. .46532
/note="LTR29 repeat: matches 571. .619 of consensus"
51039. .51094
/note="14 copies 4 mer atat 71% conserved"
51040. .51093
/note="27 copies 2 mer ta 74% conserved"
complement(51688. .52135)
/note="match: GSS: Em:AQ701298"
52348. .52817
/misc_feature      /note="match: GSS: Em:AQ568179"
52675. .53165
/note="match: GSS: Em:AQ024190"
52761. .57833
/note="LIP22 repeat: matches 1067. .6146 of consensus"
complement(61133. 61521)
/note="match: GSS: Em:AQ108129"
61134. .61471
/misc_feature      /note="match: GSS: Em:AG004108 Em:AG010234"
complement(61159. .61473)
/note="match: GSS: Em:AG004107 Em:AG010233"
complement(61162. 61459)
/misc_feature      /note="match: GSS: Em:AG004132 Em:AG010258"
61407. .61466
/note="15 copies 4 mer tata 81% conserved"
62141. .62174
/repeat_region    /note="17 copies 2 mer aa 91% conserved"
complement(62555. 63020)
/misc_feature      /note="match: GSS: Em:AQ129128"
67485. .67516
/repeat_region    /note="8 copies 4 mer taaa 90% conserved"
67741. .67901
/repeat_region    /note="LIM3c repeat: matches 12. .171 of consensus"
69516. .69979
misc_feature

Query Match      34.6%; Score 18; DB 9; Length 127029;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ATTGCCATATTGTGAGG 26
|||||
Db 103263 ATTGCCATATTGTGAGG 103280

RESULT 9
AC123468/c
LOCUS            AC123468             173249 bp    DNA    linear    HTG 23-JUL-2002
DEFINITION      Rattus norvegicus clone CH230-50L23, *** SEQUENCING IN PROGRESS
NOTE            ***, 57 unordered pieces.
ACCESSION       AC123468
VERSION         AC123468.2 GI:21908316
KEYWORDS        HTG; HTGS_PHASE1.
SOURCE          Rattus norvegicus
ORGANISM        Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 173249)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Anaratunga,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimarge,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,

```

```

Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,K., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulsegad,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Woodson,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 173249)
Worley,K.C.

Direct Submission
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 173249)
Worley,K.C.

Direct Submission
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 19, 2002 this sequence version replaced gi:21240409.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center Project name: GXPN
Center Clone name: CH230-50L23
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 124022 bases at least Q40
Consensus quality: 129548 bases at least Q30
Consensus quality: 132685 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 57 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

```
* be preserved.
* 1
* 1005: contig of 1004 bp in length
* 1104: gap of unknown length
* 2137: contig of 1033 bp in length
* 2138: gap of unknown length
* 2237: gap of unknown length
* 2238: contig of 1301 bp in length
* 3538: gap of unknown length
* 3539: contig of 1615 bp in length
* 5253: gap of unknown length
* 5254: contig of 1324 bp in length
* 5354: gap of unknown length
* 6677: gap of unknown length
* 6778: contig of 1652 bp in length
* 8430: gap of unknown length
* 8529: contig of 1497 bp in length
* 10027: gap of unknown length
* 10126: contig of 1164 bp in length
* 11290: gap of unknown length
* 11291: contig of 1500 bp in length
* 11391: gap of unknown length
* 12890: contig of 1517 bp in length
* 12891: gap of unknown length
* 14628: contig of 1638 bp in length
* 14629: gap of unknown length
* 14729: contig of 1425 bp in length
* 16154: gap of unknown length
* 16254: contig of 1517 bp in length
* 17771: gap of unknown length
* 17870: contig of 1232 bp in length
* 19102: gap of unknown length
* 19103: contig of 1137 bp in length
* 19203: gap of unknown length
* 20340: contig of 1662 bp in length
* 22101: gap of unknown length
* 22201: contig of 1926 bp in length
* 22202: gap of unknown length
* 24128: contig of 1021 bp in length
* 24228: gap of unknown length
* 25249: contig of 1957 bp in length
* 25349: gap of unknown length
* 27405: contig of 2018 bp in length
* 29423: gap of unknown length
* 29523: contig of 1229 bp in length
* 30752: gap of unknown length
* 30753: contig of 1930 bp in length
* 30853: gap of unknown length
* 32782: contig of 1930 bp in length
* 32783: gap of unknown length
* 34905: contig of 2023 bp in length
* 35005: gap of unknown length
* 36224: contig of 1219 bp in length
* 36324: gap of unknown length
* 36325: contig of 2200 bp in length
* 38525: gap of unknown length
* 38625: contig of 1179 bp in length
* 39803: gap of unknown length
* 39804: contig of 2343 bp in length
* 42246: gap of unknown length
* 42347: contig of 1171 bp in length
* 43517: gap of unknown length
* 43617: contig of 1268 bp in length
* 44885: gap of unknown length
* 44886: contig of 2165 bp in length
* 47150: gap of unknown length
* 47250: contig of 1917 bp in length
* 49167: gap of unknown length
* 49267: contig of 2878 bp in length
* 52145: gap of unknown length
* 52245: contig of 1868 bp in length
* 54113: gap of unknown length
* 54213: contig of 1860 bp in length
* 56073: gap of unknown length
* 56173: contig of 4168 bp in length
* 60341: gap of unknown length
* 60342: contig of 1690 bp in length
* 62131: gap of unknown length
* 62232: contig of 2535 bp in length
* 64766: gap of unknown length
* 64767: gap of unknown length
```

```
* 64867 68920: contig of 4054 bp in length
* 68921: gap of unknown length
* 71581: contig of 2561 bp in length
* 71582: gap of unknown length
* 71681: contig of 2916 bp in length
* 74597: gap of unknown length
* 74598: contig of 3754 bp in length
* 78451: gap of unknown length
* 78452: contig of 3830 bp in length
* 82381: gap of unknown length
* 82382: contig of 4639 bp in length
* 82482: gap of unknown length
* 87121: contig of 3529 bp in length
* 87221: gap of unknown length
* 90749: gap of unknown length
* 90849: contig of 2434 bp in length
* 90850: gap of unknown length
* 93284: contig of 4996 bp in length
* 93384: gap of unknown length
* 98380: contig of 5277 bp in length
* 98480: gap of unknown length
* 103757: contig of 4240 bp in length
* 103857: gap of unknown length
* 108097: contig of 4702 bp in length
* 108197: gap of unknown length
* 112899: contig of 5607 bp in length
* 112999: gap of unknown length
* 118606: contig of 5599 bp in length
* 118706: gap of unknown length
* 124305: contig of 5598 bp in length
* 124405: gap of unknown length
* 130003: contig of 5610 bp in length
* 130103: gap of unknown length
* 135713: contig of 7172 bp in length
* 135813: gap of unknown length
* 142985: contig of 8223 bp in length
* 143085: gap of unknown length

Query Match 34.6%; Score 18; DB 2; Length 173249;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 CAGTAGGTACTGTGCAC 51
|||||
Db 113473 CAGTAGGTACTGTGCAC 113456

RESULT 10
AC122048/c 173804 bp DNA linear HTG 27-MAY-2002
LOCUS Mus musculus chromosome UNK clone RP24-456G2, WORKING DRAFT
DEFINITION SEQUENCE, 5 unordered pieces.
ACCESSION AC122048
VERSION AC122048.1 GI:21040300
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (27-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT ----- Genome Center -----
```

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@genome.wustl.edu
----- Project Information -----
Center project name: M_BB0456G02
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171136 bases at least Q40
Consensus quality: 171554 bases at least Q30
Consensus quality: 171998 bases at least Q20
Insert size: 153000; agarose- ϕ
Insert size: 173404; sum-of-contigs
Quality coverage: 13.17 in Q20 bases; agarose- ϕ
Quality coverage: 11.69 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 35463: contig of 35463 bp in length
* 35464 35863: gap of unknown length
* 35864 76950: contig of 41387 bp in length
* 76951 77050: gap of unknown length
* 77051 139095: contig of 62045 bp in length
* 139096 139195: gap of unknown length
* 139196 150678: contig of 11483 bp in length
* 150679 150778: gap of unknown length
* 150779 173804: contig of 23026 bp in length.
*
* Location/Qualifiers
* 1..173804
* /organism="Mus musculus"
* /db_xref="taxon:10090"
* /chromosome="UNK"
* /clone="RP24-456G2"
* 1..35463
* /note="assembly_name:Contig10"
* 35564..76950
* /note="assembly_name:Contig11"
* 77051..139095
* /note="assembly_name:Contig12"
* 139196..150678
* /note="assembly_name:Contig8"
* 150779..173804
* /note="assembly_name:Contig9"
* 50438 a 36980 c 37257 g 48717 t 412 others
*
* BASE COUNT 50438 a 36980 c 37257 g 48717 t 412 others
*
* ORIGIN
*
* Query Match 34.6%; Score 18; DB 2; Length 173804;
* Best Local Similarity 100.0%; Pred. No. 1.9;
* Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
* QY 9 ATTGCCATATTTGTGAGG 26
* |||||
* Db 124356 ATTGCCATATTTGTGAGG 124339
*
* RESULT 11
* AC098609
* LOCUS AC098609 174661 bp DNA linear HTG 12-JUL-2002
* DEFINITION Rattus norvegicus clone CH230-176A4, *** SEQUENCING IN PROGRESS
* *** 75 unordered pieces.
* AC098609
* ACCESSION AC098609
* VERSION AC098609.4 GI:21729840

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
COMMENT

HTG: HTGS_PHASE1.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 174661)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Albrooks, S.L., Amaral, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbieri, J., Benton, J., Bimane, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,
Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsone, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louiseged, H.,
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapney, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogihara, G., Ogunu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G.,
Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Woodson, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 174661)
Worley, K.C.
Direct Submission
Submitted (26-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 174661)
Worley, K.C.
Direct Submission
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:17973788.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GILK
Center clone name: CH230-176A4
----- Summary Statistics -----
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329
Consensus quality: 114515 bases at least Q40
Consensus quality: 119644 bases at least Q30
Consensus quality: 123850 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 75 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1202: contig of 1202 bp in length
* 1203: gap of unknown length
* 1303: 2782: contig of 1480 bp in length
* 2883: 2882: gap of unknown length
* 2883: 4109: contig of 1227 bp in length
* 4110: 4209: gap of unknown length
* 4210: 5328: contig of 1120 bp in length
* 5330: 5429: gap of unknown length
* 5430: 6597: contig of 1168 bp in length
* 6598: 8236: contig of 1539 bp in length
* 8237: 8336: gap of unknown length
* 8337: 9717: contig of 1381 bp in length
* 9718: 9817: gap of unknown length
* 9818: 11411: contig of 1594 bp in length
* 11412: 11511: gap of unknown length
* 11512: 12717: contig of 1206 bp in length
* 12718: 12817: gap of unknown length
* 12818: 14226: contig of 1409 bp in length
* 14227: 14226: gap of unknown length
* 14327: 15519: contig of 1193 bp in length
* 15520: 15619: gap of unknown length
* 15620: 16810: contig of 1191 bp in length
* 16811: 16910: gap of unknown length
* 16911: 18166: contig of 1256 bp in length
* 18167: 18266: gap of unknown length
* 18267: 19319: contig of 1053 bp in length
* 19320: 19419: gap of unknown length
* 19420: 20989: contig of 1570 bp in length
* 20990: 21089: gap of unknown length
* 21090: 22341: contig of 1252 bp in length
* 22342: 22441: gap of unknown length
* 22442: 23883: contig of 1442 bp in length
* 23884: 23983: gap of unknown length
* 23984: 25830: contig of 1847 bp in length
* 25831: 25930: gap of unknown length
* 25931: 27517: contig of 1587 bp in length
* 27518: 27617: gap of unknown length
* 27618: 28914: contig of 1297 bp in length
* 28915: 29014: gap of unknown length
* 29015: 30444: contig of 1430 bp in length
* 30445: 30544: gap of unknown length
* 30545: 32616: contig of 2072 bp in length
* 32617: 32716: gap of unknown length
* 32717: 34119: contig of 1403 bp in length
* 34120: 34219: gap of unknown length
* 34220: 35426: contig of 1207 bp in length
* 35427: 35526: gap of unknown length
* 35527: 38132: contig of 2606 bp in length
* 38133: 38232: gap of unknown length
* 38233: 39333: contig of 1101 bp in length
* 39334: 39433: gap of unknown length
* 39434: 41013: contig of 1580 bp in length
* 41014: 41113: gap of unknown length
* 41114: 43154: contig of 2041 bp in length
* 43155: 43254: gap of unknown length
* 43255: 44711: contig of 1457 bp in length
* 44712: 44811: gap of unknown length

* 44812: 46545: contig of 1734 bp in length
* 46546: gap of unknown length
* 46646: 47888: contig of 1243 bp in length
* 47889: gap of unknown length
* 47989: 49659: contig of 1981 bp in length
* 49660: gap of unknown length
* 50089: gap of unknown length
* 50090: contig of 1390 bp in length
* 51459: gap of unknown length
* 51460: 52790: contig of 1231 bp in length
* 51560: 52890: gap of unknown length
* 52791: 54995: contig of 2105 bp in length
* 54996: gap of unknown length
* 55096: 56467: contig of 1372 bp in length
* 56468: 56567: gap of unknown length
* 56568: 58641: contig of 2074 bp in length
* 58642: 58741: gap of unknown length
* 58742: 60575: contig of 1834 bp in length
* 60576: gap of unknown length
* 60676: 61921: contig of 1246 bp in length
* 61922: gap of unknown length
* 61923: 63553: contig of 1532 bp in length
* 63554: gap of unknown length
* 63555: 65649: contig of 1996 bp in length
* 65650: gap of unknown length
* 65750: 67619: contig of 1870 bp in length
* 67620: 70164: contig of 2445 bp in length
* 70165: 70264: gap of unknown length
* 70265: 72495: contig of 2231 bp in length
* 72496: gap of unknown length
* 72596: 74134: contig of 1539 bp in length
* 74135: 74234: gap of unknown length
* 74235: 76129: contig of 1895 bp in length
* 76130: 76229: gap of unknown length
* 76230: 77645: contig of 1416 bp in length
* 77646: gap of unknown length
* 77746: 80759: contig of 3014 bp in length
* 80760: 80859: gap of unknown length
* 80860: 83194: contig of 2335 bp in length
* 83195: 83294: gap of unknown length
* 83295: 85252: contig of 1958 bp in length
* 85253: gap of unknown length
* 85254: 88126: contig of 2774 bp in length
* 88127: 88226: gap of unknown length
* 88227: 90934: contig of 2708 bp in length
* 90935: 91034: gap of unknown length
* 91035: 92932: contig of 1898 bp in length
* 92933: gap of unknown length
* 93032: 93958: contig of 2926 bp in length

Query Match 34.6%; Score 18; DB 2; Length 174661;

Best Local Similarity 100.0%; Pred. No. 1.9;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 CAGTAGGTATCTGTGCAC 51

|||||

Db 107751 CAGTAGGTATCTGTGCAC 107768

RESULT 12

AC124465/c

LOCUS

DEFINITION AC124465 175876 bp DNA linear HTG 14-JUN-2002

Mus musculus chromosome UNK clone RP24-156M16, WORKING DRAFT

SEQUENCE, 9 unordered pieces.

ACCESSION AC124465

VERSION AC124465.1 GI:21426584

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 175876)

AUTHORS McPherson, J.D. and Waterston, R.H.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

The sequence of Mus musculus clone
2 (bases 1 to 175876)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M.BB0156M16
----- Summary Statistics -----

Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Assembley program: Phrap; version 0.990319
Consensus quality: 172521 bases at least Q40
Consensus quality: 173386 bases at least Q30
Consensus quality: 174089 bases at least Q20
Insert size: 176000; agarose-fp
Quality coverage: 14.44 in Q20 bases; agarose-fp
Quality coverage: 11.81 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1051: contig of 1051 bp in length
* 1052 1151: gap of unknown length
* 1152 8912: contig of 7761 bp in length
* 8913 9012: gap of unknown length
* 9013 19813: contig of 10801 bp in length
* 19814 19913: gap of unknown length
* 19914 42142: contig of 22229 bp in length
* 42143 42242: gap of unknown length
* 42243 72359: contig of 30117 bp in length
* 72360 72459: gap of unknown length
* 72460 114685: contig of 42226 bp in length
* 114686 114785: gap of unknown length
* 114786 174506: contig of 59721 bp in length
* 174507 174606: gap of unknown length
* 174607 175470: contig of 864 bp in length
* 175471 175570: gap of unknown length
* 175571 175876: contig of 306 bp in length.

FEATURES
source

Location/Qualifiers
1..175876
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-156M16"
1..1051
/note="assembly_name:Contig36"
1152..8912
/note="assembly_name:Contig38"
9013..19813
/note="assembly_name:Contig39"
19914..42142
/note="assembly_name:Contig40"
42243..72359
/note="assembly_name:Contig41"
72460..114685
/note="assembly_name:Contig42"

misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

misc_feature 114786..174506
/note="assembly_name:Contig43"
misc_feature 174607..175470
/note="assembly_name:Contig33"
misc_feature 175571..175876
/note="assembly_name:Contig22"
BASE COUNT 46529 a 38926 c 38359 g 51247 t 815 others
ORIGIN

Query Match 34.6%; Score 18; DB 2; Length 175876;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 GAGGTCACCTGCAGTAGG 40
|||||
Db 160148 GAGGTCACCTGCAGTAGG 160131

RESULT 13
AC095423

LOCUS
DEFINITION

AC095423
Rattus norvegicus clone CH230-7C18, *** SEQUENCING IN PROGRESS ***,
69 unordered pieces.
AC095423
VERSION
HTG: HTGS_PHASE1.
KEYWORDS
SOURCE
ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 178183)

AUTHORS
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayete,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,M., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,F., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozard,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogum,M., Okwuonu,G.,
Oragunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Petrez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE

Direct Submission

Unpublished
2 (bases 1 to 178183)
Worley, K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 178183)
Worley, K.C.
Direct Submission
Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 9, 2002 this sequence version replaced gi:20975837.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GCDL
Center clone name: CH230-7C18
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 119572 bases at least Q40
Consensus quality: 124280 bases at least Q30
Consensus quality: 127521 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 89 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1036: contig of 1036 bp in length
* 1037 1136: gap of unknown length
* 1137 2639: contig of 1503 bp in length
* 2640 2739: gap of unknown length
* 2740 3821: contig of 1082 bp in length
* 3822 3921: gap of unknown length
* 3922 4977: contig of 1056 bp in length
* 4978 5077: gap of unknown length
* 5078 6101: contig of 1024 bp in length
* 6102 6201: gap of unknown length
* 6202 7216: contig of 1015 bp in length
* 7217 7315: gap of unknown length
* 7317 8320: contig of 1004 bp in length
* 8321 8420: gap of unknown length
* 8421 9504: contig of 1084 bp in length
* 9505 9604: gap of unknown length
* 9605 11209: contig of 1605 bp in length
* 11210 11309: gap of unknown length
* 11310 12920: contig of 1611 bp in length
* 12921 13020: gap of unknown length
* 13021 14274: contig of 1254 bp in length
* 14275 14374: gap of unknown length
* 14375 15499: contig of 1125 bp in length
* 15500 15599: gap of unknown length
* 15600 16609: contig of 1010 bp in length
* 16610 16709: gap of unknown length
* 16710 17853: contig of 1144 bp in length
* 17854 17953: gap of unknown length
* 17954 19487: contig of 1534 bp in length
* 19488 19588: gap of unknown length
* 19589 20644: contig of 1057 bp in length
* 20645 20744: gap of unknown length
* 20745 21970: contig of 1226 bp in length
*
* 21971 22070: gap of unknown length
* 22071 23560: contig of 1490 bp in length
* 23561 23660: gap of unknown length
* 23661 25340: contig of 1680 bp in length
* 25341 25440: gap of unknown length
* 25441 26594: contig of 1144 bp in length
* 26595 26684: gap of unknown length
* 26685 28649: contig of 1965 bp in length
* 28650 28749: gap of unknown length
* 28750 30322: contig of 1573 bp in length
* 30323 30422: gap of unknown length
* 30423 32113: contig of 1691 bp in length
* 32114 32213: gap of unknown length
* 32214 34017: contig of 1804 bp in length
* 34018 34117: gap of unknown length
* 34118 35692: contig of 1575 bp in length
* 35693 35792: gap of unknown length
* 35793 37077: contig of 1285 bp in length
* 37078 37177: gap of unknown length
* 37178 39181: contig of 2004 bp in length
* 39182 39281: gap of unknown length
* 39282 40572: contig of 1291 bp in length
* 40573 40672: gap of unknown length
* 40673 42566: contig of 1894 bp in length
* 42567 42686: gap of unknown length
* 42687 44399: contig of 1733 bp in length
* 44400 44499: gap of unknown length
* 44500 45622: contig of 1123 bp in length
* 45623 45722: gap of unknown length
* 45723 47535: contig of 1813 bp in length
* 47536 47635: gap of unknown length
* 47636 48981: contig of 1346 bp in length
* 48982 49081: gap of unknown length
* 49082 50628: contig of 1547 bp in length
* 50629 50728: gap of unknown length
* 50729 52322: contig of 1594 bp in length
* 52323 52422: gap of unknown length
* 52423 54270: contig of 1848 bp in length
* 54271 54371: gap of unknown length
* 54371 55859: contig of 1489 bp in length
* 55860 58276: gap of unknown length
* 58277 58376: contig of 2317 bp in length
* 58377 60984: contig of 2608 bp in length
* 60985 61084: gap of unknown length
* 61085 62906: contig of 1822 bp in length
* 62907 63006: gap of unknown length
* 63007 64641: contig of 1635 bp in length
* 64642 64741: gap of unknown length
* 64742 67017: contig of 2276 bp in length
* 67018 67117: gap of unknown length
* 67118 69416: contig of 2299 bp in length
* 69417 69517: gap of unknown length
* 69517 72348: contig of 2832 bp in length
* 72349 72448: gap of unknown length
* 72449 74771: contig of 2323 bp in length
* 74772 74871: gap of unknown length
* 74872 77383: contig of 2512 bp in length
* 77384 77483: gap of unknown length
* 77484 79534: contig of 2051 bp in length
* 79535 79634: gap of unknown length
* 79635 83189: contig of 3555 bp in length
* 83190 83289: gap of unknown length
* 83290 85990: contig of 2701 bp in length
* 85991 86090: gap of unknown length
* 86091 88243: contig of 2153 bp in length
* 88244 88343: gap of unknown length
* 88344 91434: contig of 3091 bp in length
* 91435 91534: gap of unknown length
* 91535 94095: contig of 2561 bp in length
* 94096 94195: gap of unknown length
* 94196 96558: contig of 2463 bp in length
* 96559 96758: gap of unknown length

```

* 96759 99395: contig of 2637 bp in length
Query Match 34.6%; Score 18; DB 2; Length 178183;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TGTGAGGTCACCTGCAGT 37
|||||
Db 62810 TGTGAGGTCACCTGCAGT 62827

RESULT 14
AC091782 195574 bp DNA linear ROD 01-JUL-2002
LOCUS Genomic sequence for Mus musculus, clone RP23-103P23, complete
sequence.
AC091782
AC091782.5 GI:21637442
HTG.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 195574)
McCombie,W.R., Spiegel,L., de la Bastide,M., Preston,R.,
Ferraro,K., Kuit,K., Nascimento,L., Zutavern,T., Balija,V.,
Bell,M., Baker,J., Miller,B., Katzenberger,F., Muller,S., King,L.,
Sullivan,P., Yang,C., Dike,S., Palmer,L., O'Shaughnessy,A. and
Dedhia,N.
Genomic sequence for Mus musculus, clone RP23-103P23, complete
sequence
Unpublished
2 (bases 1 to 195574)
McCombie,W.R.
Direct Submission
Submitted (05-JUN-2001) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
3 (bases 1 to 195574)
McCombie,W.R.
Direct Submission
Submitted (01-JUL-2002) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
On Jul 1, 2002 this sequence version replaced gi:16973717.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest.
RP23-103P23 overlaps clone RP23-349P1 (AC091785) from base 1 to
base 44451. The overlap is from base 133828 to base 178271 on
RP23-349P1.

FEATURES
    source
        Location/Qualifiers
            1..195574
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="RP23-103P23"
                /clone_lib="RPCI-23"
                2579..2590
                /note="The sequence shows one additional repeat unit
                (TTCTCCCTCC) when compared to the overlapping region on
                RP23-349P1 (AC091785)."
```

Quality coverage: 0.00 in Q20 bases; agarose-fp
Quality coverage: 7.74 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 17064: contig of 17064 bp in length
* 17065 17164: gap of unknown length
* 17165 60411: contig of 43247 bp in length
* 60412 60511: gap of unknown length
* 60512 91942: contig of 31431 bp in length
* 91943 92042: gap of unknown length
* 92043 206281: contig of 114239 bp in length
* 206282 206381: gap of unknown length
* 206382 207558: contig of 1177 bp in length.

FEATURES
Source
1. .207558
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-51605"
1. .17064
/note="assembly_name:Contig12"
clone_end:SP6
vector_side:right"
17165. .60411
/note="assembly_name:Contig14"
clone_end:T7
vector_side:left"
60512. .91942
/note="assembly_name:Contig13"
92043. .206281
/note="assembly_name:Contig15"
206382. .207558
/note="assembly_name:Contig5"
BASE COUNT 63196 a 41834 c 39615 g 62513 t 400 others
ORIGIN

Query Match 34.6%; Score 18; DB 2; Length 207558;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 CACTTGCAGTAGGTATCT 45
|||||
Db 126739 CACTTGCAGTAGGTATCT 126756

Search completed: November 26, 2002, 15:36:15
Job time : 3541 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 08:58:04 ; Search time 3142 Seconds
(without alignments)
481.651 Million cell updates/sec

Title: US-09-875-945-3

Perfect score: 52

Sequence: 1 ctctcgcattgccatattt.....gagtagtgatctgtgcaca 52

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rtd.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	52	100.0	70390	9	AC120044	Homo sapi
2	52	100.0	143409	2	AL162497	Human DNA
3	52	100.0	190740	2	AC011864	Homo sapi
4	26.8	51.5	185765	2	AC117921	Rattus no
5	26.4	50.8	147569	2	AC026312	Homo sapi
6	26.4	50.8	159535	2	AC046141	Homo sapi
7	26.4	50.8	165139	2	AC008405	Homo sapi
8	26.4	50.8	166530	9	AC069259	Homo sapi
9	26.4	50.8	173978	2	AC020649	Homo sapi
10	26.4	50.8	178340	2	AC022120	Homo sapi
11	26.4	50.8	193252	2	AC008658	Homo sapi
12	25.6	49.2	47349	9	AL732363	Human DNA
13	25.6	49.2	175621	2	AC053486	Homo sapi
14	25.6	49.2	185896	2	AC069564	Mus muscu
15	25.6	49.2	219180	2	AC092251	Mus muscu
16	25.2	48.5	5806	10	MUSBCL22	L31532 Mus musculu
17	25.2	48.5	169620	2	AC012674	Homo sapi
18	25.2	48.5	263190	2	AC015657	Mus muscu
19	25	48.1	179583	2	AC011931	Homo sapi
20	25	48.1	179616	9	AC078994	Homo sapi
21	25	48.1	247475	2	AC008247	Homo sapi
22	24.8	47.7	207378	9	AC009062	Homo sapi
23	24.4	46.9	82452	10	AL663054	Mouse DNA
24	24.4	46.9	120027	2	AC020981	Homo sapi
25	24.4	46.9	154468	2	AC018409	Homo sapi
26	24.4	46.9	159438	2	AC016621	Homo sapi
27	24.4	46.9	164139	2	AC110476	Rattus no
28	24.4	46.9	178563	9	AL357833	Human DNA
29	24.4	46.9	180029	2	AC013756	Homo sapi
30	24.4	46.9	188570	9	AC022113	Homo sapi
31	24.4	46.9	193849	2	AC084829	Mus muscu
32	24.4	46.9	196472	2	AC011862	Homo sapi
33	24.2	46.5	76089	2	AC026945	Homo sapi
34	24.2	46.5	184557	2	AC112619	Rattus no
35	24.2	46.5	196555	2	AC092882	Homo sapi
36	24.2	46.5	201437	2	AC015981	Homo sapi
37	24	46.2	108443	2	AC126150	Rattus no
38	24	46.2	151310	2	AC110332	Rattus no
39	24	46.2	151310	2	AC110332	Rattus no
40	24	46.2	179210	2	AC118839	Rattus no
41	24	46.2	195003	2	AC112812	Rattus no
42	24	46.2	206214	2	AC095934	Rattus no
43	24	46.2	227634	2	AC111203	Rattus no
44	23.8	45.8	1466	8	AY084871	Arabidops
45	23.8	45.8	16202	8	AT283321	A.thaliana

ALIGNMENTS

RESULT 1

AC120044

LOCUS

DEFINITION

Homo sapiens chromosome 15 clone RP11-163G3 map 15, LOW-PASS

SEQUENCE SAMPLING.

ACCESSION

AC120044

VERSION

AC120044.2

KEYWORDS

HTG; HTGS_PHASE0.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 70390)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 15, clone RP11-163G3

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 70390)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faró,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL
REFERENCE
AUTHORS
Direct Submission
Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 70390)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faró,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Lakocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., MacDonald,P., Major,J., Marquis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 6, 2002 this sequence version replaced gi:20389713.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26755
Center clone name: 163_G_3

* NOTE: This record contains 87 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 675: contig of 675 bp in length
676 775: gap of 100 bp
776 1474: contig of 699 bp in length
1475 1574: gap of 100 bp
1575 2306: contig of 732 bp in length
2307 2406: gap of 100 bp
2407 3138: contig of 722 bp in length
3129 3228: gap of 100 bp
3229 3934: contig of 706 bp in length
3935 4034: gap of 100 bp
4035 4737: contig of 703 bp in length
4738 4837: gap of 100 bp
4838 5537: contig of 700 bp in length
5538 5637: gap of 100 bp
5638 6348: contig of 711 bp in length
6349 7145: contig of 697 bp in length
7146 7245: gap of 100 bp
7246 7963: contig of 718 bp in length
7964 8063: gap of 100 bp
8064 8793: contig of 730 bp in length
8794 8893: gap of 100 bp
8894 9616: contig of 723 bp in length
9617 9716: gap of 100 bp
9717 10420: contig of 704 bp in length
10421 10520: gap of 100 bp
10521 11223: contig of 709 bp in length
11230 11329: gap of 100 bp
11330 12028: contig of 699 bp in length
12029 12128: gap of 100 bp
12129 12842: contig of 714 bp in length
12843 12942: gap of 100 bp
12943 13659: contig of 717 bp in length
13660 13759: gap of 100 bp
13760 14468: contig of 709 bp in length
14469 14568: gap of 100 bp
14569 15286: contig of 718 bp in length
15287 15386: gap of 100 bp
15387 16109: contig of 723 bp in length
16110 16209: gap of 100 bp
16210 16923: contig of 714 bp in length
16924 17023: gap of 100 bp
17024 17742: contig of 719 bp in length
17743 17842: gap of 100 bp
17843 18564: contig of 722 bp in length
18565 18664: gap of 100 bp
18665 19386: contig of 722 bp in length
19387 19486: gap of 100 bp
19487 20186: contig of 700 bp in length
20187 20286: gap of 100 bp
20287 20993: contig of 707 bp in length
20994 21093: gap of 100 bp
21094 21798: contig of 705 bp in length
21799 21898: gap of 100 bp
21899 22600: contig of 702 bp in length
22601 22700: gap of 100 bp
22701 23389: contig of 689 bp in length
23390 23489: gap of 100 bp
23490 24205: contig of 716 bp in length
24206 24305: gap of 100 bp
24306 25022: contig of 717 bp in length
25023 25122: gap of 100 bp
25123 25830: contig of 708 bp in length
25831 25930: gap of 100 bp
25931 26634: contig of 704 bp in length
26635 26734: gap of 100 bp
26735 27442: contig of 708 bp in length
27443 27542: gap of 100 bp
27543 28262: contig of 720 bp in length

```
* 28263 28362: gap of 100 bp
* 28363 29077: contig of 715 bp in length
* 29078 29177: gap of 100 bp
* 29178 29878: contig of 701 bp in length
* 29879 29978: gap of 100 bp
* 29979 30682: contig of 704 bp in length
* 30683 30782: gap of 100 bp
* 30783 31483: contig of 703 bp in length
* 31486 31585: gap of 100 bp
* 31586 32292: contig of 707 bp in length
* 32293 32392: gap of 100 bp
* 32393 33098: contig of 706 bp in length
* 33099 33198: gap of 100 bp
* 33199 33901: contig of 703 bp in length
* 33902 34001: gap of 100 bp
* 34002 34714: contig of 713 bp in length
* 34715 34814: gap of 100 bp
* 34815 35538: contig of 724 bp in length
* 35539 35638: gap of 100 bp
* 35639 36348: contig of 710 bp in length
* 36349 36448: gap of 100 bp
* 36449 37152: contig of 704 bp in length
* 37153 37252: gap of 100 bp
* 37253 37983: contig of 733 bp in length
* 37986 38085: gap of 100 bp
* 38086 38791: contig of 706 bp in length
* 38792 38891: gap of 100 bp
* 38892 39598: contig of 707 bp in length
* 39599 39698: gap of 100 bp
* 39699 40408: contig of 708 bp in length
* 40407 40506: gap of 100 bp
* 40507 41222: contig of 716 bp in length
* 41223 41322: gap of 100 bp
* 41323 42037: contig of 715 bp in length
* 42038 42137: gap of 100 bp
* 42138 42841: contig of 704 bp in length
* 42842 42941: gap of 100 bp
* 42942 43666: contig of 725 bp in length
* 43667 43766: gap of 100 bp
* 43767 44483: contig of 717 bp in length
* 44484 44583: gap of 100 bp

Query Match 100.0%; Score 52; DB 2; Length 70390;
Best Local Similarity 100.0%; Pred. No. 1.8e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 CTCCTCGCATTCGCATATTTGTGAGGTGACCTGCGTAGGTATCGTCACA 52
Db 58456 CTCCTCGCATTCGCATATTTGTGAGGTGACCTGCGTAGGTATCGTCACA 68507

RESULT 2
AL162497/c
LOCUS AL162497 143409 bp DNA linear PRI 06-JUN-2001
DEFINITION Human DNA sequence from clone RP11-313L9 on chromosome 13, complete
sequence.
ACCESSION AL162497
VERSION AL162497.20 GI:14329908
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143409)
AUTHORS Mashreghi-Mohammadi,M.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 8, 2001 this sequence version replaced gi:14280409.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
```

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/Projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr13> RP11-313L9 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-313L9 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP11-313L9 is at 143409 in this sequence. The true left end of clone RP11-40E6 is at 100074 in this sequence. The true right end of clone RP11-358F13 is at 100 in this sequence.

FEATURES

	Location/Qualifiers
source	1. 143409 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="13" /clone_lib="RP11-313L9" /clone_lib="RPCI-11.2"
repeat_region	1. 350 /note="L1MA9 repeat: matches 5421. .5789 of consensus"
repeat_region	351. 654 /note="L1PBI repeat: matches 5846. .6155 of consensus"
repeat_region	655. .957 /note="L1MA9 repeat: matches 5789. .6085 of consensus"
repeat_region	2061. .2106 /note="23 copies 2 mer tg 91% conserved"
repeat_region	2063. .2106 /note="11 copies 4 mer tggg 93% conserved"
repeat_region	2138. .2422 /note="L1MD3 repeat: matches 6691. .7023 of consensus"
repeat_region	2422. .3051 /note="L1MC4 repeat: matches 7168. .7837 of consensus"
repeat_region	3674. 4545 /note="L1PA4 repeat: matches 5272. .6144 of consensus"
repeat_region	5701. .5910 /note="7 copies 30 mer 89% conserved"
repeat_region	6490. .6613 /note="MIR repeat: matches 92. .225 of consensus"
repeat_region	9599. 9648 /note="25 copies 2 mer aa 94% conserved"
repeat_region	10251. .10559 /note="Alusx repeat: matches 2. .312 of consensus"
repeat_region	11134. .11521 /note="L1MB7 repeat: matches 5762. .6161 of consensus"
repeat_region	12667. .12710 /note="11 copies 4 mer tggg 81% conserved"
repeat_region	13210. .13309 /note="L2 repeat: matches 2559. .2674 of consensus"
repeat_region	13794. .14148 /note="THEJC repeat: matches 1. .371 of consensus"
repeat_region	14816. .15030 /note="MIR repeat: matches 49. .256 of consensus"
repeat_region	14982. .15037 /note="L2 repeat: matches 2651. .2705 of consensus"

repeat_region	15476..15940	/note="L2 repeat: matches 1314. .1835 of consensus"	repeat_region	/note="13 copies 49 mer 62% conserved"
repeat_region	16305..16609	/note="AluX repeat: matches 1. .300 of consensus"	repeat_region	/note="11 copies 52 mer 62% conserved"
repeat_region	16683..16722	/note="20 copies 2 mer tt 77% conserved"	repeat_region	/note="205 copies 2 mer aa 56% conserved"
repeat_region	16948..17073	/note="MLT1C repeat: matches 348. .464 of consensus"	repeat_region	/note="15 copies 20 mer 60% conserved"
repeat_region	17074..17360	/note="AluJ repeat: matches 13. .299 of consensus"	repeat_region	/note="53 copies 4 mer taaa 59% conserved"
repeat_region	17361..17661	/note="MLT1C repeat: matches 7. .348 of consensus"	repeat_region	/note="64 copies 3 mer ata 60% conserved"
repeat_region	17974..18347	/note="L1MD1 repeat: matches 5659. .6029 of consensus"	repeat_region	/note="3 copies 30 mer 78% conserved"
repeat_region	18399..18554	/note="3 copies 52 mer 75% conserved"	repeat_region	/note="L1ME1 repeat: matches 5550. .6144 of consensus"
repeat_region	18400..18549	/note="5 copies 30 mer 73% conserved"	repeat_region	/note="L2 repeat: matches 2054. .2250 of consensus"
repeat_region	18413..18552	/note="5 copies 28 mer 74% conserved"	repeat_region	/note="L2 repeat: matches 1446. .1832 of consensus"
repeat_region	18565..18738	/note="L1MD2 repeat: matches 6150. .6335 of consensus"	repeat_region	/note="MIR repeat: matches 48. .147 of consensus"
repeat_region	19070..19105	/note="9 copies 4 mer gata 88% conserved"	repeat_region	/note="20 copies 2 mer tt 82% conserved"
repeat_region	19363..19664	/note="L2 repeat: matches 2436. .2750 of consensus"	repeat_region	/note="7 copies 5 mer tgttt 85% conserved"
repeat_region	20032..20053	/note="11 copies 2 mer aa 100% conserved"	repeat_region	/note="L2 repeat: matches 2699. .2747 of consensus"
repeat_region	21272..21554	/note="AluSq repeat: matches 1. .283 of consensus"	repeat_region	/note="THE1B repeat: matches 1. .364 of consensus"
repeat_region	22754..22918	/note="3 copies 55 mer 77% conserved"	repeat_region	/note="L2 repeat: matches 2629. .2750 of consensus"
repeat_region	26601..26684	/note="L2 repeat: matches 2668. .2748 of consensus"	repeat_region	/note="L2 repeat: matches 2259. .2489 of consensus"
repeat_region	26885..26814	/note="AluJb repeat: matches 2. .131 of consensus"	repeat_region	45491..45781
repeat_region	26815..27110	/note="AluSq repeat: matches 1. .303 of consensus"	Query Match	100.0%; Score 52; DB 9; Length 143409;
repeat_region	27111..27285	/note="AluJb repeat: matches 130. .307 of consensus"	Best Local Similarity	100.0%; Pred No. 1.8e-10;
repeat_region	27286..27578	/note="L2 repeat: matches 2359. .2668 of consensus"	Matches 52; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
repeat_region	29945..30257	/note="AluY repeat: matches 1. .306 of consensus"	QY	1 CTCTCGCATTCATATTTGTGAGTGCACCTGTCAGTAGGTATCTGTGCACA 52
repeat_region	30812..31078	/note="AluX repeat: matches 1. .267 of consensus"	Db	135485 CTCTCGCATTCATATTTGTGAGTGCACCTGTCAGTAGGTATCTGTGCACA 135434
repeat_region	31887..32091	/note="MER58A repeat: matches 1. .222 of consensus"	RESULT 3	
repeat_region	32818..32853	/note="9 copies 4 mer acac 100% conserved"	AC011864	AC011864
repeat_region	32856..32957	/note="MER97C repeat: matches 185. .289 of consensus"	LOCUS	AC011864.3 GI:7230118
repeat_region	32958..33406	/note="MLT1C repeat: matches 1. .466 of consensus"	DEFINITION	Homo sapiens clone RP11-16C4, WORKING DRAFT SEQUENCE, 51 unordered pieces.
repeat_region	33407..33867	/note="MER97C repeat: matches 289. .726 of consensus"	ACCESSION	AC011864
repeat_region	33895..34718	/note="8 copies 103 mer 61% conserved"	VERSION	AC011864.3
repeat_region	33915..34519	/note="121 copies 5 mer atata 56% conserved"	KEYWORDS	HTG; HTGS-PHASE1; HTGS-DRAFT.
repeat_region	33920..34103	/note="92 copies 2 mer at 57% conserved"	SOURCE	Homo sapiens.
repeat_region	33924..34553	/note="21 copies 30 mer 61% conserved"	ORGANISM	Homo sapiens
repeat_region	33925..34730	/note="13 copies 62 mer 61% conserved"	REFERENCE	1 (bases 1 to 190740)
repeat_region	33940..34709	/note="14 copies 55 mer 61% conserved"	AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
repeat_region	33954..34213	/note="13 copies 20 mer 70% conserved"	TITLE	Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
repeat_region	33999..34698	/note="25 copies 28 mer 58% conserved"	JOURNAL	Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
repeat_region	34014..34650		REFERENCE	Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
			AUTHORS	Farreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
				Galagan,J., Gardyna,S., Grant,G., Hags,B., Hearford,A., Horton,L.,
				Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
				Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
				McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meidtm,J.,
				Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
				Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
				Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

```

Tsefaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE
JOURNAL
COMMENT
Submitted (15-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6479010.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBK
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3517
Center clone name: 16_C4
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 121051 bases at least Q40
Consensus quality: 152215 bases at least Q30
Consensus quality: 171265 bases at least Q20
Insert size: 167000; agarose-fp
Quality coverage: 3.5 in Q20 bases; agarose-fp
Quality coverage: 3.1 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 51 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1197: contig of 1197 bp in length
* 1198 1297: gap of 100 bp
* 1298 2601: contig of 1304 bp in length
* 2602 2701: gap of 100 bp
* 2702 3906: contig of 1205 bp in length
* 3907 4006: gap of 100 bp
* 4007 5023: contig of 1017 bp in length
* 5024 5123: gap of 100 bp
* 5124 6559: contig of 1436 bp in length
* 6560 6659: gap of 100 bp
* 6660 8178: contig of 1519 bp in length
* 8179 8278: gap of 100 bp
* 8279 9354: contig of 1076 bp in length
* 9355 9454: gap of 100 bp
* 9455 10758: contig of 1304 bp in length
* 10759 10858: gap of 100 bp
* 10859 12011: contig of 1153 bp in length
* 12012 12111: gap of 100 bp
* 12112 13508: contig of 1397 bp in length
* 13509 13608: gap of 100 bp
* 13609 14887: contig of 1279 bp in length
* 14888 14987: gap of 100 bp
* 14988 16156: contig of 1169 bp in length
* 16157 16256: gap of 100 bp
* 16257 17785: contig of 1529 bp in length
* 17786 17885: gap of 100 bp
* 17886 18947: contig of 1062 bp in length
* 18948 19047: gap of 100 bp
* 19048 20330: contig of 1283 bp in length
* 20331 20430: gap of 100 bp
* 20431 21914: contig of 1484 bp in length
* 21915 22014: gap of 100 bp
* 22015 23101: contig of 1087 bp in length
* 23102 23201: gap of 100 bp
* 23202 24447: contig of 1246 bp in length
* 24448 24547: gap of 100 bp
*
* 24548 26032: contig of 1485 bp in length
* 26033 26132: gap of 100 bp
* 26133 28012: contig of 1880 bp in length
* 28013 28112: gap of 100 bp
* 28113 30280: contig of 2168 bp in length
* 30281 30380: gap of 100 bp
* 30381 31436: contig of 1056 bp in length
* 31437 31536: gap of 100 bp
* 31537 33906: contig of 2370 bp in length
* 33907 34006: gap of 100 bp
* 34007 35444: contig of 1438 bp in length
* 35445 35544: gap of 100 bp
* 35545 38284: contig of 2740 bp in length
* 38285 38384: gap of 100 bp
* 38385 40159: contig of 1775 bp in length
* 40160 40259: gap of 100 bp
* 40260 42089: contig of 1830 bp in length
* 42090 42189: gap of 100 bp
* 42190 43894: contig of 1705 bp in length
* 43895 43994: gap of 100 bp
* 43995 46499: contig of 2505 bp in length
* 46500 46599: gap of 100 bp
* 46600 48033: contig of 1434 bp in length
* 48034 48133: gap of 100 bp
* 48134 51777: contig of 3644 bp in length
* 51778 51877: gap of 100 bp
* 51878 55116: contig of 3239 bp in length
* 55117 55216: gap of 100 bp
* 55217 57721: contig of 2505 bp in length
* 57722 57821: gap of 100 bp
* 57822 61349: contig of 3528 bp in length
* 61350 61449: gap of 100 bp
* 61450 65218: contig of 3769 bp in length
* 65219 65318: gap of 100 bp
* 65319 68743: contig of 3425 bp in length
* 68744 68843: gap of 100 bp
* 68844 74044: contig of 5201 bp in length
* 74045 74144: gap of 100 bp
* 74145 78742: contig of 4598 bp in length
* 78743 78842: gap of 100 bp
* 78843 82807: contig of 3965 bp in length
* 82808 82907: gap of 100 bp
* 82908 87683: contig of 4776 bp in length
* 87684 87783: gap of 100 bp
* 87784 92706: contig of 4923 bp in length
* 92707 92806: gap of 100 bp
* 92807 99946: contig of 7140 bp in length
* 99947 100046: gap of 100 bp
* 100047 107272: contig of 7226 bp in length
* 107273 107372: gap of 100 bp
* 107373 115832: contig of 8460 bp in length
* 115833 115932: gap of 100 bp
* 115933 124872: contig of 8940 bp in length
* 124873 124972: gap of 100 bp
* 124973 134820: contig of 9848 bp in length
* 134821 134920: gap of 100 bp
* 134921 143495: contig of 8575 bp in length
* 143496 143595: gap of 100 bp
* 143596 152927: contig of 9332 bp in length
* 152928 153027: gap of 100 bp
* 153028 164076: contig of 11049 bp in length
* 164077 164176: gap of 100 bp
* 164177 173869: contig of 9693 bp in length
* 173870 173969: gap of 100 bp
* 173970 190740: contig of 16771 bp in length.
*
* Location/Qualifiers
* 1..190740
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /clone="RP11-16C4"
* /clone_lib="RPC1-11 Human Male BAC"
* 1..1197
* /note="assembly_fragment"
*
FEATURES
Source
misc_feature
1..1197
/note="assembly_fragment"

```

[illegible]

* 8899 10353: contig of 1455 bp in length
* 10354 10453: gap of unknown length
* 11983 11983: contig of 1530 bp in length
* 11984 12083: gap of unknown length
* 12084 13228: contig of 1145 bp in length
* 13229 13328: gap of unknown length
* 13329 14682: contig of 1354 bp in length
* 14683 14783: gap of unknown length
* 14783 16008: contig of 1226 bp in length
* 16009 16108: gap of unknown length
* 16109 17203: contig of 1095 bp in length
* 17204 17303: gap of unknown length
* 17304 18699: contig of 1396 bp in length
* 18700 18799: gap of unknown length
* 18800 20460: contig of 1661 bp in length
* 20461 20560: gap of unknown length
* 20561 21719: contig of 1159 bp in length
* 21720 21819: gap of unknown length
* 21820 23225: contig of 1406 bp in length
* 23226 23326: gap of unknown length
* 23326 24854: contig of 1529 bp in length
* 24855 24954: gap of unknown length
* 24955 26171: contig of 1217 bp in length
* 26172 26272: gap of unknown length
* 26272 28758: contig of 2487 bp in length
* 28759 28858: gap of unknown length
* 28859 30130: contig of 1272 bp in length
* 30131 30230: gap of unknown length
* 30231 31799: contig of 1569 bp in length
* 31800 31899: gap of unknown length
* 31900 33592: contig of 1693 bp in length
* 33593 33692: gap of unknown length
* 33693 35047: contig of 1355 bp in length
* 35048 35147: gap of unknown length
* 35148 37283: contig of 2136 bp in length
* 37284 37383: gap of unknown length
* 37384 39165: contig of 1782 bp in length
* 39166 39265: gap of unknown length
* 39266 41184: contig of 1919 bp in length
* 41185 41284: gap of unknown length
* 41285 43006: contig of 1722 bp in length
* 43007 43106: gap of unknown length
* 43107 44551: contig of 1445 bp in length
* 44552 44651: gap of unknown length
* 44652 46694: contig of 2043 bp in length
* 46695 46794: gap of unknown length
* 46795 48306: contig of 1512 bp in length
* 48307 48406: gap of unknown length
* 48407 50972: contig of 2566 bp in length
* 50973 52968: contig of 1896 bp in length
* 52969 53068: gap of unknown length
* 53069 54760: contig of 1692 bp in length
* 54761 54860: gap of unknown length
* 54861 56972: contig of 2112 bp in length
* 56973 57072: gap of unknown length
* 57073 58717: contig of 1645 bp in length
* 58718 58817: gap of unknown length
* 58818 60551: contig of 1734 bp in length
* 60552 60651: gap of unknown length
* 60652 63129: contig of 2478 bp in length
* 63130 63229: gap of unknown length
* 63230 65097: contig of 1868 bp in length
* 65098 65197: gap of unknown length
* 65198 68734: contig of 3537 bp in length
* 68735 68834: gap of unknown length
* 68835 71067: contig of 2233 bp in length
* 71068 71167: gap of unknown length
* 71168 73017: contig of 1850 bp in length
* 73017 73117: gap of unknown length
* 73118 76139: contig of 3022 bp in length
* 76140 76239: gap of unknown length
* 76240 78677: contig of 2438 bp in length

* 78678 78777: gap of unknown length
* 78778 80149: contig of 1372 bp in length
* 80150 80249: gap of unknown length
* 80250 82961: contig of 2712 bp in length
* 82962 83061: gap of unknown length
* 83062 84794: contig of 1733 bp in length
* 84795 84894: gap of unknown length
* 84895 87616: contig of 2722 bp in length
* 87617 87716: gap of unknown length
* 87717 89199: contig of 1483 bp in length
* 89200 89299: gap of unknown length
* 89300 92252: contig of 2953 bp in length
* 92253 92352: gap of unknown length
* 92353 96303: contig of 3951 bp in length
* 96304 96403: gap of unknown length
* 96404 100348: contig of 3945 bp in length
* 100349 100448: gap of unknown length
* 100449 102780: contig of 2332 bp in length
* 102781 102880: gap of unknown length
* 102881 105070: contig of 2190 bp in length

Query Match 51.5%; Score 26.8; DB 2; Length 185765;

Best Local Similarity 73.9%; Pred. No. 4.1;

Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 7 GCATTGCCATATTGTGAGGTCACCTGTCAGTAGTATCTGTGCACA 52

Db 111193 GCATTTTATATGTCAGGTCAGTGGGTAGGCATATGTACACA 111238

RESULT 5

AC026312

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC026312 147569 bp DNA linear HTG 03-MAY-2002
Homo sapiens chromosome 3 clone RP11-151A21, WORKING DRAFT
SEQUENCE, 1 unordered piece.
AC026312
AC026312.18 GI:20428724
HTG: HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 147569)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amarantunge, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbaria, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleaveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorelli, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Huly, J., Huly, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,

Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scher, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, F., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 147569)
 Worley, K.C.
 Direct Submission
 Submitted (22-MAR-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 147569)
 Worley, K.C.
 Direct Submission
 Submitted (03-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On May 3, 2002 this sequence version replaced gi:20335524.

 Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

 Project Information
 Center project name: HAPF
 Center clone name: RP11-151A21

 Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-primer Bodipy: 5% of reads
 Chemistry: Dye-terminator Big Dye: 95% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 157145 bases at least Q40
 Consensus quality: 159796 bases at least Q30
 Consensus quality: 161774 bases at least Q20
 Estimated insert size: 151039; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see <http://www.hgsc.bcm.tmc.edu/docs/genbank/draft.data.html>).
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 1 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 *
 1 147569: contig of 147569 bp in length.
 Location/Qualifiers
 1..147569
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-151A21"
 BASE COUNT 43888 c 28495 c 30409 g 44777 t
 ORIGIN
 Query Match 50.8%; Score 26.4; DB 2; Length 147569;
 Best Local Similarity 75.0%; Pred. No. 5.9;
 Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 QY 1 CTCCTCGCATTCGATATTGTGAGCTCACTTCAGTAGGTATC 44
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 77672 CTCCTCCTATTGCTATATGTTGAGCTCAAGTGTAGGAAGTATC 77715

RESULT 6
 AC046141
 LOCUS
 DEFINITION
 Homo sapiens chromosome 5 clone RP11-93E8, WORKING DRAFT SEQUENCE,
 25 unordered pieces.
 AC046141 AC0244223
 VERSION
 AC046141.8 GI:20335583
 KEYWORDS
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 159535)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
 Alsbrooks, J.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,
 Barbaria, S., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flag, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Homs, F., Howard, S., Huber, J., Huly, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louissegh, H.,
 Lozato, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scher, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 159535)
 Worley, K.C.
 Direct Submission
 Submitted (13-APR-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 159535)
 Worley, K.C.
 Direct Submission
 Submitted (04-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Apr 28, 2002 this sequence version replaced gi:16117923.

 Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>


```

* 5840      8362: contig of 2523 bp in length
* 8363      8462: gap of unknown length
* 8463      10917: contig of 2455 bp in length
* 10918     11017: gap of unknown length
* 11018     12465: contig of 1448 bp in length
* 12466     12565: gap of unknown length
* 12566     17665: contig of 5100 bp in length
* 17666     17765: gap of unknown length
* 17766     21349: contig of 3584 bp in length
* 21350     21449: gap of unknown length
* 21450     27140: contig of 5691 bp in length
* 27141     27240: gap of unknown length
* 27241     35996: contig of 8756 bp in length
* 35997     36096: gap of unknown length
* 36097     48933: contig of 12837 bp in length
* 48934     49033: gap of unknown length
* 49034     70704: contig of 21671 bp in length
* 70705     70804: gap of unknown length
* 70805     95784: contig of 24980 bp in length
* 95785     95884: gap of unknown length
* 95885     165139: contig of 69255 bp in length.

FEATURES             Location/Qualifiers
     source            1..165139
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTC-27603"
                     /clone_lib="CalTech human BAC library C"
BASE COUNT           50118 a 30980 c 30784 g 51856 t 1401 others
ORIGIN

Query Match          50.8%; Score 26.4; DB 2; Length 165139;
Best Local Similarity 69.2%; Pred. No. 6;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY      1  CTCCTCGCATGCCCATATTGTCAGGTCAGTGCAGTAGGTCATCGTCGACGACA 52
        |||| ||||| || |||| | || |||| | || || || || || || || ||
Db 80678 CTCGAGCATGCCCATGTCGTGAATGCTCCAGTTCGAGATCCCTCGGACA 80627

RESULT 8
AC069259/c
LOCUS          AC069259             166530 bp      DNA      linear      PRI 02-APR-2002
DEFINITION     Homo sapiens 3 BAC RP11-163H6 (Roswell Park Cancer Institute Human
BAC Library) complete sequence.
ACCESSION      AC069259
VERSION        AC069259.14  GI:19774265
KEYWORDS       HTG.
SOURCE         Homo sapiens.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 166530)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Blinage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,M., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homs,I.F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,

```

Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louised,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstein,G., and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 166530)
Worley,K.C.
Direct Submission
Submitted (23-MAY-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 166530)
Worley,K.C.
Direct Submission
Submitted (26-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 166530)
Worley,K.C.
Direct Submission
Submitted (28-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 166530)
Worley,K.C.
Direct Submission
Submitted (02-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Mar 28, 2002 this sequence version replaced gi:19172574.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect <1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2

reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES Location/Qualifiers

1..166530
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-163H6"

repeat_region complement(218..517)
repeat_region /rpt_family="AluSx"
970..1045
repeat_region /rpt_family="MER5B"
1143..1193
repeat_region /rpt_family="MER5B"
1389..1423
repeat_region /rpt_family="AT-rich"
1432..1739
repeat_region /rpt_family="AluSp"
complement(1835..1902)
repeat_region /rpt_family="MER47"

repeat_region 3159..3566
repeat_region /rpt_family="MSTB"
3565..3684
STS /standard_name="7145"

repeat_region 3590..3634
repeat_region /rpt_family="(TTTA)n"
complement(3635..3913)
repeat_region /rpt_family="AluSc"
4013..4033
repeat_region /rpt_family="AT-rich"
4418..4789
repeat_region /rpt_family="MLT1A1"
complement(4817..4901)
repeat_region /rpt_family="L2"
complement(5362..5586)
repeat_region /rpt_family="MIR"
complement(5681..5989)
repeat_region /rpt_family="AluSx"
complement(6465..6541)
repeat_region /rpt_family="MIR"
7586..7733
repeat_region /rpt_family="FRAM"

repeat_region complement(9882..10168)
repeat_region /rpt_family="AluJo"
10541..10831
repeat_region /rpt_family="AluY"
11399..11513
repeat_region /rpt_family="MIR"
12238..12296
repeat_region /rpt_family="(CA)n"

repeat_region 13493..13581
repeat_region /rpt_family="MLT1F"
13712..13783
repeat_region /rpt_family="MLT1A1"
14588..14609
repeat_region /rpt_family="AT-rich"

repeat_region 14846..15124
repeat_region /rpt_family="AluSx"
15229..15448
repeat_region /rpt_family="L2"
15573..15644
repeat_region /rpt_family="L2"

repeat_region complement(16132..16428)
repeat_region /rpt_family="AluSp"
17525..17556
repeat_region /rpt_family="CT-rich"
18271..18565
repeat_region /rpt_family="AluJb"
19586..19872
repeat_region /rpt_family="AluSx"
complement(20010..20159)
repeat_region /rpt_family="MIR"
22543..22606
repeat_region /rpt_family="MIR"
23155..23449
repeat_region /rpt_family="AluSx"
23723..23743
repeat_region /rpt_family="AT-rich"
24099..24150
repeat_region /rpt_family="AluJ/monomer"
24204..24490
repeat_region /rpt_family="AluJo"
25677..25795
repeat_region /rpt_family="FLAM_C"
26394..26554

Query Match 50.8%; Score 26.4; DB 9; Length 166530;
Best Local Similarity 75.0%; Pred. No. 6;
Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 CTCCTCGCATTCATATTTGGAGGCTACTTGCACGTAGGTATC 44
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 117266 CTCCTCTATTGCTATATGCTGAGCTCAAGTGTAGGAAGTATC 117223

RESULT 9
AC020649/c

LOCUS AC020649 Homo sapiens chromosome 3 clone RP11-163H6, WORKING DRAFT SEQUENCE,
DEFINITION 12 unordered pieces.

ACCESSION AC020649.9 GI:17352394
VERSION AC020649
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE HTG; HTGS_PHASE1; HTGS_DRAFT.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 173978)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaral, H.C., Are, J.R., Banks, T., Barbara, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, B., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flag, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulvik, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulsegue, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenkwo, S.,

Ogum, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I., Sodergren, E., Soudaie, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Vallalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 173978)
Worley, K.C. Submission
Submitted (08-JAN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 5, 2001 this sequence version replaced gi:9929569.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMXG
Center clone name: RP11-163H6
----- Summary Statistics
Sequencing vector: Plasmid; M77789
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 174642 bases at least Q40
Consensus quality: 187023 bases at least Q30
Consensus quality: 193456 bases at least Q20
Estimated insert size: 187358; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 4.5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 74914: contig of 74914 bp in length
* 74915 75014: gap of unknown length
* 75015 100184: contig of 25170 bp in length
* 100185 100284: gap of unknown length
* 100285 121498: contig of 21214 bp in length
* 121499 121598: gap of unknown length
* 121599 134584: contig of 12986 bp in length
* 134585 134684: gap of unknown length
* 134685 148013: contig of 13329 bp in length
* 148014 148113: gap of unknown length
* 148114 155389: contig of 7276 bp in length
* 155390 155489: gap of unknown length
* 155490 159060: contig of 3471 bp in length
* 159061 159060: gap of unknown length
* 159061 162043: contig of 2983 bp in length
* 162044 162143: gap of unknown length
* 162144 165299: contig of 3156 bp in length
* 165300 165399: gap of unknown length
* 165400 168513: contig of 3114 bp in length
* 168514 168613: gap of unknown length
* 168614 171097: contig of 2484 bp in length
* 171098 171197: gap of unknown length

FEATURES
source
* 171198 173978: contig of 2781 bp in length.
Location/Qualifiers
1. 173978
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-163H6"
BASE COUNT 53093 a 35559 c 32862 g 51347 t 1117 others
ORIGIN
Query Match 50.8% Score 26.4; DB 2; Length 173978;
Best Local Similarity 75.0%; Pred. No. 6;
Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 1 CTCCTCGCATTCATATTTGTGAGGTCACATTCAGTAGGATC 44
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 74676 CTCCTCTATTGCTATAATGCTGAGCTCAAGTGTAGGAAGTATC 74633
RESULT 10
AC022120/c
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CTC-570I13, WORKING DRAFT SEQUENCE,
19 unordered pieces.
AC022120 178340 bp DNA linear HTG 20-APR-2001
VERSION AC022120.5 GI:12830145
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 178340)
DOE Joint Genome Institute.
Direct Submission
Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Feb 14, 2001 this sequence version replaced gi:7711697.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 404265
Center clone name: CIT-HSPC_570I13

Summary Statistics
Consensus quality: 162107 bases at least Q40
Consensus quality: 169630 bases at least Q30
Consensus quality: 172125 bases at least Q20
Estimated insert size: 175000; pulse field gel estimation
Estimated insert size: 176540; sum-of-contigs estimation
Quality coverage: 4.93 in Q20 bases; pulse field gel estimation
Quality coverage: 4.88 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1122: contig of 1122 bp in length
* 1123 1222: gap of unknown length
* 1223 1233: contig of 1159 bp in length
* 1234 2381: gap of unknown length
* 2382 2481: gap of unknown length
* 2482 3335: contig of 1054 bp in length
* 3336 3635: gap of unknown length
* 3636 4777: contig of 1142 bp in length

* 4778 4877: gap of unknown length
* 4878 6017: contig of 1140 bp in length
* 6018 6117: gap of unknown length
* 6118 8064: contig of 1947 bp in length
* 8065 8164: gap of unknown length
* 8165 10855: contig of 2691 bp in length
* 10856 10955: gap of unknown length
* 10956 12098: contig of 1143 bp in length
* 12099 12198: gap of unknown length
* 12199 14071: contig of 1873 bp in length
* 14072 14171: gap of unknown length
* 14172 18044: contig of 3873 bp in length
* 18045 18144: gap of unknown length
* 18145 21874: contig of 3730 bp in length
* 21875 21974: gap of unknown length
* 21975 24474: contig of 2500 bp in length
* 24475 24574: gap of unknown length
* 24575 37909: contig of 13335 bp in length
* 37910 38009: gap of unknown length
* 38010 52447: contig of 14438 bp in length
* 52448 52547: gap of unknown length
* 52548 69563: contig of 17016 bp in length
* 69564 69664: gap of unknown length
* 69664 92561: contig of 22898 bp in length
* 92562 92661: gap of unknown length
* 92662 113850: contig of 21189 bp in length
* 113851 113950: gap of unknown length
* 113951 137315: contig of 23365 bp in length
* 137316 137415: gap of unknown length
* 137416 178340: contig of 40925 bp in length.

FEATURES

source

1. .178340
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CFC-570I13"
/clone_lib="Caltech human BAC library C"
BASE COUNT 53260 a 33379 c 33069 g 56805 t 1827 others
ORIGIN

Query Match 50.8%; Score 26.4; DB 2; Length 178340;
Best Local Similarity 69.2%; Pred. No. 6;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 CTCCTCGCATTCGCATATTGTGAGTGACTTGCAGTAGTATCTGTGCACA 52

Db 33590 CTCACGCGATTCGCACATGTGTGAATTCCTCCACTTGAGATCCCTGGACA 33539

RESULT 11
AC008658/c

LOCUS Homo sapiens chromosome 5 clone CFB-22H6, WORKING DRAFT SEQUENCE,
25 ordered pieces.
DEFINITION AC008658 193252 bp DNA linear HTG 04-OCT-2001

AC008658.5 GI:14579686

HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 193252)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 5

Unpublished

2 (bases 1 to 193252)

DOE Joint Genome Institute.

Direct Submission

Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

On Jul 1, 2001 this sequence version replaced gi:12830084.

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 72826, H396
Center clone name: C1978SKB_22H6

Summary Statistics
Consensus quality: 175357 bases at least Q40
Consensus quality: 185996 bases at least Q30
Consensus quality: 188955 bases at least Q20
Estimated insert size: 210000; pulse field gel estimation
Quality coverage: 8.21 in Q20 bases; sum-of-contigs estimation
Quality coverage: 9.03 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1 6260: contig of 6260 bp in length
* 6261 6360: gap of unknown length
* 6361 27077: contig of 20717 bp in length
* 27078 27177: gap of unknown length
* 27178 32402: contig of 5225 bp in length
* 32403 32502: gap of unknown length
* 32503 62904: contig of 30402 bp in length
* 62905 73192: contig of 10188 bp in length
* 73193 73292: gap of unknown length
* 73293 89434: contig of 16142 bp in length
* 89435 89534: gap of unknown length
* 89535 92100: contig of 2566 bp in length
* 92101 92200: gap of unknown length
* 92201 94868: contig of 2668 bp in length
* 94869 94968: gap of unknown length
* 94969 96171: contig of 1203 bp in length
* 96172 96271: gap of unknown length
* 96272 97625: gap of unknown length
* 97626 97725: contig of 3451 bp in length
* 97726 101275: gap of unknown length
* 101276 101789: contig of 6414 bp in length
* 101790 107789: gap of unknown length
* 107790 109602: contig of 1813 bp in length
* 109603 109702: gap of unknown length
* 109703 111861: contig of 2159 bp in length
* 111862 111961: gap of unknown length
* 111962 115292: contig of 3331 bp in length
* 115293 115392: gap of unknown length
* 115393 123026: contig of 7634 bp in length
* 123027 126439: contig of 3313 bp in length
* 126440 126539: gap of unknown length
* 126540 129405: contig of 2866 bp in length
* 129406 129505: gap of unknown length
* 129506 133646: contig of 4141 bp in length
* 133647 133746: gap of unknown length
* 133747 136412: contig of 2666 bp in length
* 136413 136512: gap of unknown length
* 136513 153606: contig of 17094 bp in length
* 153607 153706: gap of unknown length
* 153707 176401: contig of 22695 bp in length
* 176402 176501: gap of unknown length
* 176502 179132: contig of 2631 bp in length
* 179133 179232: gap of unknown length
* 179233 188289: contig of 9057 bp in length
* 188290 188389: gap of unknown length
* 188390 193252: contig of 4863 bp in length.

FEATURES	SOURCE
----------	--------

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 14:37:15 ; Search time 83 Seconds
(without alignments)
276.026 Million cell updates/sec

Title: US-09-875-945-3
Perfect score: 52
Sequence: 1 ctctctgcattgcatattt.....gcagtagtgatctgtgcaca 52

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 288240 seqs, 220289906 residues

Word size : 0
Total number of hits satisfying chosen parameters: 576480

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_NA_New.*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	15	28.8	581	5	US-09-513-999C-13496
C 2	15	28.8	74962	6	US-10-274-974-3
C 3	14	26.9	380	5	US-09-513-999C-33243
C 4	14	26.9	411	6	US-10-266-131-2791
C 5	14	26.9	5024	6	US-10-240-965-129
C 6	14	26.9	177851	1	PCT-US02-33542-3
C 7	13	25.0	36	6	US-10-298-148-38
C 8	13	25.0	112	5	US-09-513-999C-33851
C 9	13	25.0	149	5	US-09-513-999C-10529
C 10	13	25.0	210	5	US-09-513-999C-23279
C 11	13	25.0	212	5	US-09-513-999C-25433
C 12	13	25.0	328	5	US-09-513-999C-18166
C 13	13	25.0	341	5	US-09-513-999C-36478
C 14	13	25.0	420	5	US-09-513-999C-35989
C 15	13	25.0	491	6	US-10-266-131-2388
C 16	13	25.0	514	5	US-09-513-999C-8234
C 17	13	25.0	585	5	US-09-620-607B-1442
C 18	13	25.0	592	5	US-09-724-676-18377
C 19	13	25.0	627	6	US-10-273-006-46
C 20	13	25.0	636	5	US-09-513-999C-3963
C 21	13	25.0	728	5	US-09-513-999C-10372
C 22	13	25.0	847	5	US-09-724-676-40385
C 23	13	25.0	917	5	US-09-724-676-40377
C 24	13	25.0	918	5	US-09-724-676-40349
C 25	13	25.0	962	5	US-09-724-676-15881
C 26	13	25.0	993	6	US-10-092-411A-1165

ALIGNMENTS

RESULT 1

US-09-513-999C-13496/c
; Sequence 13496, Application US/09513999C
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; SOFTWARE: Patent.pm
; NUMBER OF SEQ ID NOS: 36681
; SEQ ID NO 13496
; LENGTH: 581
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 379
; OTHER INFORMATION: s-g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 381
; OTHER INFORMATION: y=c or t
US-09-513-999C-13496

Query Match 28.8%; Score 15; DB 5; Length 581;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 AGGTCACCTGTCAGTA 38

|||||

Db 86 AGGTCACCTGTCAGTA 72

RESULT 2

US-10-274-974-3/c
; Sequence 3, Application US/10274974
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; FILE REFERENCE: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CLO00871 DIV
; CURRENT APPLICATION NUMBER: US/10/274,974
; CURRENT FILING DATE: 2002-10-18

```
; PRIOR APPLICATION NUMBER: 09/685,853
; PRIOR FILING DATE: 2000-10-11
; PRIOR APPLICATION NUMBER: 60/182,194
; PRIOR FILING DATE: 2000-02-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 74962
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(74962)
; OTHER INFORMATION: n = A,T,C or G
US-10-274-974-3

Query Match      28.8%; Score 15; DB 6; Length 74962;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TCCTGCGCATTCGCAT 16
Db 479 TCCTGCGCATTCGCAT 465

RESULT 3
US-09-513-999C-33243
; Sequence 33243, Application US/09513999C
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 33243
; LENGTH: 380
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-33243

Query Match      26.9%; Score 14; DB 5; Length 380;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 TAGGTATCTGTGCA 50
Db 356 TAGGTATCTGTGCA 369

RESULT 4
US-10-266-131-2791
; Sequence 2791, Application US/10266131
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Novel Murine Polynucleotide Tags and
; FILE REFERENCE: LEX-0030-USA
; CURRENT APPLICATION NUMBER: US/10/266,131
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: US/09/617,675
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/143,878
; PRIOR FILING DATE: 1999-07-15
; NUMBER OF SEQ ID NOS: 2908
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 2791
; LENGTH: 411
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(411)
; OTHER INFORMATION: n = A,T,C or G
US-10-266-131-2791

Query Match      26.9%; Score 14; DB 6; Length 411;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TGCCATATTTGTGA 24
Db 269 TGCCATATTTGTGA 282

RESULT 5
US-10-240-965-129
; Sequence 129, Application US/10240965
; GENERAL INFORMATION:
; APPLICANT: INCITE GENOMICS, INC.
; APPLICANT: SHIFFMAN, Dov
; APPLICANT: SOMOGYI, Roland
; APPLICANT: LAWN, Richard M.
; APPLICANT: SEILHAMER, Jeffrey J.
; APPLICANT: PORTER, Gordon J.
; APPLICANT: MIKITA, Thomas
; TITLE OF INVENTION: GENES EXPRESSED IN FOAM CELL DIFFERENTIATION
; FILE REFERENCE: PA-0025 PCT
; CURRENT APPLICATION NUMBER: US/10/240,965
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: 60/195,106
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 276
; SOFTWARE: PERL Program
; SEQ ID NO 129
; LENGTH: 5024
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 898945.14
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 2112-2457
; OTHER INFORMATION: a, t, c, g, or other
US-10-240-965-129

Query Match      26.9%; Score 14; DB 6; Length 5024;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CATTGCCATATTTG 21
Db 1598 CATTGCCATATTTG 1611

RESULT 6
PCT-US02-33542-3
; Sequence 3, Application PC/TUS0233542
; GENERAL INFORMATION:
; APPLICANT: Exelixis, Inc.
; TITLE OF INVENTION: MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-120
; CURRENT APPLICATION NUMBER: PCT/US02/33542
; CURRENT FILING DATE: 2002-10-21
; PRIOR APPLICATION NUMBER: 60/338,733
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/357,600
```

; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 177851
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-33542-3

Query Match 26.9%; Score 14; DB 1; Length 177851;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TGCATATTGTGA 24
|||||

Db 128832 TGCATATTGTGA 128845

RESULT 7
US-10-298-148-38/c
; Sequence 38, Application US/10298148
; GENERAL INFORMATION:
; APPLICANT: Cox III, George N
; TITLE OF INVENTION: Derivatives of Growth Hormone and Related Proteins
; FILE REFERENCE: 4152-1-PUS
; CURRENT APPLICATION NUMBER: US/10/298,148
; PRIOR FILING DATE: 2002-11-15
; PRIOR APPLICATION NUMBER: US/09/462,941
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: 60/052,516
; PRIOR FILING DATE: 1997-07-14
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 36
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PCR Primer
US-10-298-148-38

Query Match 25.0%; Score 13; DB 6; Length 36;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 ACTTGCAGTAGGT 41
|||||

Db 31 ACTTGCAGTAGGT 19

RESULT 8
US-09-513-999C-33851
; Sequence 33851, Application US/09513999C
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 33851
; LENGTH: 112
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-33851

Query Match 25.0%; Score 13; DB 5; Length 112;

Best Local Similarity 100.0%; Pred. No. 99;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 TGCAGTAGGTATC 44
|||||

Db 95 TGCAGTAGGTATC 107

RESULT 9
US-09-513-999C-10529
; Sequence 10529, Application US/09513999C
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 10529
; LENGTH: 149
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-10529

Query Match 25.0%; Score 13; DB 5; Length 149;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ATTGTGAGGTCA 29
|||||

Db 88 ATTGTGAGGTCA 100

RESULT 10
US-09-513-999C-23279/c
; Sequence 23279, Application US/09513999C
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 23279
; LENGTH: 210
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 146
; OTHER INFORMATION: w=a or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 192
; OTHER INFORMATION: b=c or g or t
US-09-513-999C-23279

Query Match 25.0%; Score 13; DB 5; Length 210;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 ATATTTGTGAGGT 27
|||||

Db 59 ATATTGTGAGGT 47

RESULT 11

US-09-513-999C-25433

; Sequence 25433, Application US/09513999C

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Duclert, A.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

; FILE REFERENCE: 59.US2.REG

; CURRENT APPLICATION NUMBER: US/09/513.999C

; CURRENT FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/122,487

; PRIOR FILING DATE: 1999-02-26

; NUMBER OF SEQ ID NOS: 36681

; SOFTWARE: Patent.pm

; SEQ ID NO 25433

; LENGTH: 212

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-513-999C-25433

Query Match

Best Local Similarity 25.0%; Score 13; DB 5; Length 212;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 TTTGTGAGGTCAC 30

|||||

Db 52 TTTGTGAGGTCAC 64

RESULT 12

US-09-513-999C-18166

; Sequence 18166, Application US/09513999C

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Duclert, A.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

; FILE REFERENCE: 59.US2.REG

; CURRENT APPLICATION NUMBER: US/09/513.999C

; CURRENT FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/122,487

; PRIOR FILING DATE: 1999-02-26

; NUMBER OF SEQ ID NOS: 36681

; SOFTWARE: Patent.pm

; SEQ ID NO 18166

; LENGTH: 328

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 317

; OTHER INFORMATION: v-a or c or g

US-09-513-999C-18166

Query Match

Best Local Similarity 25.0%; Score 13; DB 5; Length 328;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 CATTGCCATATT 20

|||||

Db 143 CATTGCCATATT 155

RESULT 13

US-09-513-999C-36478/c

; Sequence 36478, Application US/09513999C

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Duclert, A.

```
; LOCATION: 379
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 400
; OTHER INFORMATION: n=a, g, c or t
US-09-513-999C-35989

Query Match      25.0%; Score 13; DB 5; Length 420;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 28 CACTTGCACTAGG 40
    |
Db 214 CACTTGCACTAGG 226
    |

RESULT 15
US-10-266-131-2388/c
; Sequence 2388, Application US/10266131
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Novel Murine Polynucleotide Tags and
; TITLE OF INVENTION: Mutant Cells and Mutant Animals Defined Thereby
; FILE REFERENCE: LEX-0030-USA
; CURRENT APPLICATION NUMBER: US/10/266,131
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: US/09/617,675
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/143,878
; PRIOR FILING DATE: 1999-07-15
; NUMBER OF SEQ ID NOS: 2908
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2388
; LENGTH: 491
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(491)
; OTHER INFORMATION: n = A,T,C or G
US-10-266-131-2388

Query Match      25.0%; Score 13; DB 6; Length 491;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 TTGCCATATTGT 22
    |
Db 235 TTGCCATATTGT 223
    |

Search completed: November 26, 2002, 20:13:16
Job time : 132 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 08:58:00 ; Search time 295 Seconds

(without alignments)
396.962 Million cell updates/sec

Title: US-09-875-945-3

Perfect score: 52

Sequence: 1 ctctcgcattgccattt.....gcagtagtatctgtgcaca 52

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	100.0	52	AA140729	Human IRS-2 gene r
2	23.8	45.8	1466	AAC36478	Arabidopsis thalia
3	23.4	45.0	560	ABA61825	Human foetal liver
4	23.4	45.0	560	ABA29408	Probe #7874 for ge
5	23.4	45.0	560	AAK10139	Human brain expres
6	23.4	45.0	560	AAK36034	Human bone marrow
7	23.4	45.0	560	AAI41749	Probe #10435 used
8	23.4	45.0	560	ABS10207	Human genome-deriv
9	23.2	44.6	3899	ABL26712	Drosophila melanog

c	10	22.6	43.5	3446	23	ABL06375	Drosophila melanog
	11	22.6	43.5	4875	23	ABL06362	Drosophila melanog
	12	22.6	43.5	12692	23	ABL06374	Drosophila melanog
	13	22.4	43.1	1543	21	AAC44807	Zea mays DNA fragm
c	14	22.4	43.1	92584	21	AAF22288	BAC containing rep
c	15	22	42.3	1611	23	AAS65418	cDNA encoding nove
c	16	21.8	41.9	482	23	ABK42057	Human bone marrow
c	17	21.8	41.9	3258	23	AAS96245	Feline herpes viru
c	18	21.8	41.9	10803	18	AAT69857	Human immune/haema
	19	21.8	41.9	23241	22	AAK84291	Human immune/haema
	20	21.8	41.9	23241	22	AAK87225	Human cardiovascu
c	21	21.8	41.9	26048	22	AAS36056	Fusarium venenatum
	22	21.6	41.5	209	21	AAF11174	Human polynucleoti
c	23	21.6	41.5	1981	22	AAI58978	Listeria innocua D
	24	21.6	41.5	3011208	24	ABQ69245	Human breast cance
	25	21.4	41.2	396	22	AAI17351	Human breast cance
	26	21.4	41.2	787	22	AAF22542	Nucleotide sequenc
	27	21.2	40.8	264	19	AAV29274	Drosophila melanog
c	28	21.2	40.8	664	23	ABL09949	Pinus radiata tran
c	29	21.2	40.8	690	21	AAC56269	Mouse haematopoiet
	30	21.2	40.8	1935	24	RAD38777	Drosophila melanog
	31	21.2	40.8	3368	23	ABL09948	Drosophila melanog
	32	21.2	40.8	3629	23	ABL22374	Drosophila melanog
c	33	21.2	40.8	6354	23	ABL07560	Drosophila melanog
	34	21.2	40.8	10126	22	ABAI8955	Human nervous syst
c	35	21.2	40.8	10126	22	ABAI9362	Human nervous syst
	36	21.2	40.8	1830121	17	AAT42063	Haemophilus influe
c	37	21	40.4	244	22	ABAI4561	Human nervous syst
c	38	21	40.4	244	22	ABAI4562	Human nervous syst
c	39	21	40.4	244	22	ABAI4563	Human nervous syst
	40	21	40.4	973	21	AAC98967	Human pancreatic c
	41	21	40.4	3410	23	ABL18618	Drosophila melanog
c	42	21	40.4	4809	23	ABL14992	Drosophila melanog
	43	21	40.4	4983	22	RAF41629	Novel protein kina
c	44	21	40.4	16125	20	AAZ21226	Human CGICE genomi
c	45	20.8	40.0	150	19	AAI12205	Human biallelic po

ALIGNMENTS

RESULT 1
AA140729
ID AAL40729 standard; DNA; 52 BP.
XX
AC AAL40729;
XX
DT 25-SEP-2002 (first entry)
XX
DE Human IRS-2 gene related oligonucleotide 5956, SEQ ID No 3.
XX
KW Antidiabetic; anorectic; human insulin regulated substance; IRS-2 gene;
XX diabetes; obesity; metabolic disorder; ds.
OS Homo sapiens.
XX
PN WO200194410-A1.
XX
PD 13-DEC-2001.
XX
PF 08-JUN-2001; 2001WO-SE01308.
XX
PR 08-JUN-2000; 2000US-210207P.
PR 09-JUN-2000; 2000SE-0002189.
XX
(METC-) METCON MEDICIN AB.
XX Smith U;
XX WPI; 2002-566443/60.
XX New non-coding sequences isolated upstream of human insulin regulated
PT substance-2 gene, useful as marker for predicting, diagnosing and

treating metabolic diseases or disorders e.g. obesity and diabetes -
Claim 1; Fig 5; 38pp; English.

The invention relates to an isolated, substantially purified nucleotide sequence which is a non-coding sequence upstream of human insulin regulated substance (IRS-2) gene. The isolated IRS-2 gene nucleic acid region or its homologue is useful as a marker for insulin regulating action in an assay for evaluating or screening substances for insulin regulating properties in vitro, where adipocytes, hepatic cells, muscle tissue cells or pancreatic cells are used as model cells. The isolated IRS-2 gene nucleic acid region, its homologue, an IRS-2 transcript, or a substance information derived from an IRS-2 transcript, or a substance having insulin regulated properties is useful for manufacture of a medicament for treating diabetes and obesity, and as a veterinary preparation. The isolated IRS-2 gene nucleic acid region is also useful for diagnosis of IRS-2 related metabolic disorders, diabetes and/or for differentiating between various types or stages of the disorders, and for determining if a patient in need of treatment with an insulin regulating substance has the predisposition to respond to the treatment, comprising measuring activation of IRS-2 by determining the amount or relative increase/decrease of the IRS-2 protein or corresponding mRNA, where the isolated IRS-2 gene nucleic acid region is used as a marker when administering IRS to a sample of cells such as blood, adipocyte, muscle or liver cells taken from the patient. This polynucleotide sequence represents an isolated nucleic acid which is a non-coding sequence upstream of the human insulin regulated substance (IRS-2) gene of the invention.

Sequence 52 BP; 10 A; 13 C; 12 G; 17 T; 0 other;

Query Match 100.0%; Score 52; DB 24; Length 52;
Best Local Similarity 100.0%; Pred. No. 2.6e-11;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CTCCTCGATTGCCATATTTGTAGGTCACCTGCAGTAGGTATCTGTGCACA 52
|||||
Db 1 CTCCTCGATTGCCATATTTGTAGGTCACCTGCAGTAGGTATCTGTGCACA 52

RESULT 2
AAC36478/c
ID AAC36478 standard; DNA; 1466 BP.

XX AAC36478;
AC
XX
DT 17-Oct-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 13950.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.

21-APR-1999; 99US-0130449.
23-APR-1999; 99US-0130510.
23-APR-1999; 99US-0130891.
28-APR-1999; 99US-0131449.
30-APR-1999; 99US-0132048.
30-APR-1999; 99US-0132407.
30-APR-1999; 99US-0132484.
04-MAY-1999; 99US-0132485.
05-MAY-1999; 99US-0132486.
06-MAY-1999; 99US-0132487.
07-MAY-1999; 99US-0132863.
11-MAY-1999; 99US-0134256.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
14-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
20-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
21-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.

[illegible]

AAK36034
ID AAK36034 standard; DNA; 560 BP.
XX
AC AAK36034;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 10591.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00668.
XX
PF 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
DR Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 10591; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 560 BP; 186 A; 110 C; 101 G; 162 T; 1 other;
Query Match 45.0%; Score 23.4; DB 22; Length 560;
Best Local Similarity 73.2%; Pred. No. 13;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
XX
DE 17-OCT-2001 (first entry)
XX
XX Probe #10435 used to measure gene expression in human placenta sample.
XX
PF Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.

AAK36034
ID AAK36034 standard; DNA; 560 BP.
XX
AC AAK36034;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 10591.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00668.
XX
PF 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
DR Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 10591; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 560 BP; 186 A; 110 C; 101 G; 162 T; 1 other;
Query Match 45.0%; Score 23.4; DB 22; Length 560;
Best Local Similarity 73.2%; Pred. No. 13;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
XX
DE 17-OCT-2001 (first entry)
XX
XX Probe #10435 used to measure gene expression in human placenta sample.
XX
PF Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.

AAK36034
ID AAK36034 standard; DNA; 560 BP.
XX
AC AAK36034;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 10591.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00668.
XX
PF 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
DR Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 10435; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 560 BP; 186 A; 110 C; 101 G; 162 T; 1 other;
Query Match 45.0%; Score 23.4; DB 22; Length 560;
Best Local Similarity 73.2%; Pred. No. 13;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
XX
DE 17-OCT-2001 (first entry)
XX
XX Probe #10435 used to measure gene expression in human placenta sample.
XX
PF Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.

XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID No 10435; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 560 BP; 186 A; 110 C; 101 G; 162 T; 1 other;
Query Match 45.0%; Score 23.4; DB 22; Length 560;
Best Local Similarity 73.2%; Pred. No. 13;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
XX
DE 17-OCT-2001 (first entry)
XX
XX Probe #10435 used to measure gene expression in human placenta sample.
XX
PF Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00665.
XX
XX 04-FEB-2000; 2000US-180312P.
XX
XX 26-MAY-2000; 2000US-207456P.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID No 10435; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 560 BP; 186 A; 110 C; 101 G; 162 T; 1 other;
Query Match 45.0%; Score 23.4; DB 22; Length 560;
Best Local Similarity 73.2%; Pred. No. 13;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
XX
DE 17-OCT-2001 (first entry)
XX
XX Probe #10435 used to measure gene expression in human placenta sample.
XX
PF Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00665.
XX
XX 04-FEB-2000; 2000US-180312P.
XX
XX 26-MAY-2000; 2000US-207456P.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID No 10435; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 560 BP; 186 A; 110 C; 101 G; 162 T; 1 other;
Query Match 45.0%; Score 23.4; DB 22; Length 560;
Best Local Similarity 73.2%; Pred. No. 13;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
XX
DE 17-OCT-2001 (first entry)
XX
XX Probe #10435 used to measure gene expression in human placenta sample.
XX
PF Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.

PT interactions -

XX Claim 1; SEQ ID NO 13568; 21pp + Sequence Listing; English.

XX

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 4875 BP; 1266 A; 1110 C; 1108 G; 1391 T; 0 other;

Query Match 43.5%; Score 22.6; DB 23; Length 4875;

Best Local Similarity 75.7%; Pred. NO. 45;

Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0

QY 15 ATATTGTCAGGTCACCTGCAGTAGGTATCTGTGCAC 51

||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 4547 ATATATGGATGCCGTTTGGCAATATGTCGTCTGTGCAC 4583

RESULT 12

ABL06374

ID ABL06374 standard; cDNA; 12692 BP.

XX

AC ABL06374;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13604.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

XX WPI; 2001-656860/75.

DR P-FSDB; ABB62271.

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

XX Claim 1; SEQ ID NO 13604; 21pp + Sequence Listing; English.

XX

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 4875 BP; 1266 A; 1110 C; 1108 G; 1391 T; 0 other;

Query Match 43.5%; Score 22.6; DB 23; Length 4875;

Best Local Similarity 75.7%; Pred. NO. 45;

Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0

QY 15 ATATTGTCAGGTCACCTGCAGTAGGTATCTGTGCAC 51

||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 4547 ATATATGGATGCCGTTTGGCAATATGTCGTCTGTGCAC 4583

RESULT 12

ABL06374

ID ABL06374 standard; cDNA; 12692 BP.

XX

AC ABL06374;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13604.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

XX WPI; 2001-656860/75.

DR P-FSDB; ABB62271.

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

XX Claim 1; SEQ ID NO 13604; 21pp + Sequence Listing; English.

XX

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 4875 BP; 1266 A; 1110 C; 1108 G; 1391 T; 0 other;

Query Match 43.5%; Score 22.6; DB 23; Length 4875;

Best Local Similarity 75.7%; Pred. NO. 45;

Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0

QY 15 ATATTGTCAGGTCACCTGCAGTAGGTATCTGTGCAC 51

||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 4547 ATATATGGATGCCGTTTGGCAATATGTCGTCTGTGCAC 4583

RESULT 12

ABL06374

ID ABL06374 standard; cDNA; 12692 BP.

XX

AC ABL06374;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13604.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

XX WPI; 2001-656860/75.

DR P-FSDB; ABB62271.

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

XX Claim 1; SEQ ID NO 13604; 21pp + Sequence Listing; English.

XX

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 4875 BP; 1266 A; 1110 C; 1108 G; 1391 T; 0 other;

Query Match 43.5%; Score 22.6; DB 23; Length 4875;

Best Local Similarity 75.7%; Pred. NO. 45;

Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0

QY 15 ATATTGTCAGGTCACCTGCAGTAGGTATCTGTGCAC 51

||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 4547 ATATATGGATGCCGTTTGGCAATATGTCGTCTGTGCAC 4583

RESULT 12

ABL06374

ID ABL06374 standard; cDNA; 12692 BP.

XX

AC ABL06374;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13604.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

XX WPI; 2001-656860/75.

DR P-FSDB; ABB62271.

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

XX Claim 1; SEQ ID NO 13604; 21pp + Sequence Listing; English.

XX

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149923.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	24-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	23-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match

Best Local Similarity

Score 22.4; DB 21; Length 1543;
Pred. No. 41;

	Matches	26;	Conservative	0;	Mismatches	6;	Indels	0;	Gaps	0;
Oy	3	CCTGCATTGCCATATTTGTGAGTCACTTCG	34							
Db	1420	CCCCGATATTCACATTTTTTGAGTGCCTTC	1451							
RESULT 14										
AAAF22288/C	ID	AAF22288 standard; DNA; 92584 BP.								
XX	AC	AAF22288;								
XX	DT	20-MAR-2001 (first entry)								
XX	DE	BAC containing repeats from centromeres 1-4 #11.								
XX	KW	Centromere; microsome; vector; ds.								
XX	OS	Arabidopsis thaliana.								
XX	PN	WO200055325-A2.								
XX	PD	21-SEP-2000.								
XX	PF	17-MAR-2000; 2000WO-US07392.								
XX	PR	18-MAR-1999; 99US-0125219.								
XX	PR	01-APR-1999; 99US-0127409.								
XX	PR	18-MAY-1999; 99US-0134770.								
XX	PR	13-SEP-1999; 99US-0153584.								
XX	PR	17-SEP-1999; 99US-0154603.								
XX	PA	(UYCH-) UNIV CHICAGO.								
XX	PI	Preuss D, Copenhagen G, Keith K;								
XX	DR	WPI; 2000-587529/55.								
XX	PT	Recombinant DNA construct comprising a plant centromere, useful for producing stably inherited microsomes which can serve as vectors for the construction of transgenic plant and animal cells -								
XX	PS	Claim 102; Page 508-529; 1449pp; English.								
XX	CC	The present invention relates to a recombinant DNA construct of a plant (Arabidopsis thaliana) centromere. The constructs are useful for producing stably inherited microsomes which can serve as vectors for the construction of transgenic plant and animal cells expressing selected proteins such as hormones, enzymes, interleukins, clotting factors, cytokines, antibodies, and growth factors.								
SQ	Sequence	92584 BP; 27840 A; 18113 C; 19835 G; 26796 T; 0 other;								
	Query Match	43.1%; Score 22.4; DB 21; Length 92584;								
	Best Local Similarity	66.7%; Pred. No. 1.le+02;								
	Matches	32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;								
Oy	2	TCTCGCATTCGCCATATTTGTGAGTCACTTCGATAGGTATCTGTGC	49							
Db	19981	TCTTAGCGTTGCCACGTGAGTGACATCACATGCGGACGACATGTGTAC	19934							
RESULT 15										
AAS65418/C	ID	AAS65418 standard; cDNA; 1611 BP.								
XX	AC	AAS65418;								
XX	DT	13-FEB-2002 (first entry)								
XX	DE	DNA encoding novel human diagnostic protein #1222.								

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX
OS Homo sapiens.

XX
PN WO200175067-A2.

XX
PD
11-OCT-2001.

XX
PF
30-MAR-2001; 2001WO-US08631.

XX
PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.
XX

PA (HYSE-) HYSEQ INC.
XX

PI Drmanac RT, Liu C, Tang YT;
XX

DR WPI; 2001-639362/73.
DR P-PSDB; ABG01231.

XX
PT New isolated poly

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX
PS Claim 1: SEQ ID No 1222; 103pp; English.

xx The invention relates to isolated polynucleotide (I) and
cc polypeptide (II) sequences. (I) is useful as hybridisation probes,
cc polymerase chain reaction (PCR) primers, oligomers, and for chromosome
cc and gene mapping, and in recombinant production of (II). The
cc polynucleotides are also used in diagnostics as expressed sequence tags
cc for identifying expressed genes. (I) is useful in gene therapy techniques
cc to restore normal activity of (II) or to treat disease states involving
cc (II). (II) is useful for generating antibodies against it, detecting or
cc quantitating a polypeptide in tissue, as molecular weight markers and as
cc a food supplement. (II) and its binding partners are useful in medical
cc imaging of sites expressing (II). (I) and (II) are useful for treating
cc disorders involving aberrant protein expression or biological actions
cc The polypeptide and polynucleotide sequences have applications in
cc diagnostics, forensics, gene mapping, identification of mutations
cc responsible for genetic disorders or other traits to assess biodiversity
cc and to produce other types of data and products dependent on DNA and
cc amino acid sequences. AAS64197-AA594564 represent novel human
cc diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at <ftp.wipo.int/pub/published> pct sequences.

Sequence 1611 BP: 365 A: 473 C: 476 G: 297 T: 0 other;
XX
SO

Query Match	42.3%	Score 22;	DB 23;	Length 1611;
Best Local Similarity	67.4%	Pred. NO. 60;		
Matches	31.	Conservative	0.	Mismatches
			15.	Indels
			0.	Gaps

ov 7 GCATTGCCATATTTGTGAGGTCACCTTGCAGTAGGTATCTGTGCACA 52

db 1270 GGATTGCGATATTAAATGAGGGCCTCGTTCAGTCCAGATCTTTCCAAA 1225

Search completed: November 26, 2002, 10:11:58
Job time : 319 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 14:33:10 ; Search time 3784 Seconds
(without alignments)
345.510 Million cell updates/sec

Title: US-09-875-945-3
Perfect score: 52
Sequence: 1 ctctctcgattgcatattt.....gcagtaggtatctgtgcaca 52

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 24791104 seqs, 12571243825 residues

Word size : 0
Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_NA_Main.*
1: /cgn2_6/ptodata/1/pna/US06_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US08_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US081_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US082_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US083_COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US084_COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US085_COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US086_COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US087_COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US088_COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US089_COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US090_COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US091_COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US092_COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US093_COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US094_COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US095A_COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US095B_COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US095C_COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US095D_COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US096A_COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US096B_COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US096C_COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US096D_COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US096E_COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US097A_COMB.seq.*
29: /cgn2_6/ptodata/1/pna/US097B_COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US097C_COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US098A_COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US098B_COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US098C_COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US099A_COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US099B_COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US099C_COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US099D_COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US100A_COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US100B_COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US101A_COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US101B_COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US102A_COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US102B_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	52	100.0	52	33	US-09-875-945-3
2	18	34.6	276	25	US-09-654-617-303468
3	18	34.6	276	27	US-09-684-016-303468
4	17	32.7	2797	34	US-09-912-935-37
5	16	30.8	184	5	US-08-196-363-1858
6	16	30.8	184	5	US-08-196-363A-1858
7	16	30.8	184	5	US-08-196-363-1858
8	16	30.8	184	32	US-09-859-490-1858
9	16	30.8	184	33	US-09-866-555-4699
10	16	30.8	225	27	US-09-698-014-3804
11	16	30.8	319	56	US-60-129-391-156
12	16	30.8	328	17	US-09-304-517A-269440
13	16	30.8	328	17	US-09-371-146A-268282
14	16	30.8	328	22	US-09-565-240-18654
15	16	30.8	328	37	US-09-985-678-269440
16	16	30.8	335	16	US-09-289-768-33675
17	16	30.8	335	35	US-09-939-397-33675
18	16	30.8	346	23	US-09-606-977-17133
19	16	30.8	368	23	US-09-606-977-47510
20	16	30.8	390	17	US-09-304-517A-252559
21	16	30.8	390	17	US-09-391-630-17026
					Sequence 3, Appli
					Sequence 303468,
					Sequence 303468,
					Sequence 37, Appl
					Sequence 1858, Ap
					Sequence 1858, Ap
					Sequence 1858, Ap
					Sequence 4699, Ap
					Sequence 3804, Ap
					Sequence 156, App
					Sequence 269440,
					Sequence 268282,
					Sequence 18654, A
					Sequence 269440,
					Sequence 33675, A
					Sequence 33675, A
					Sequence 71733, A
					Sequence 47510, A
					Sequence 252559,
					Sequence 17026, A

RESULT 5

US-08-196-363-1858/c

; Sequence 1858, Application US/08196363

; GENERAL INFORMATION:

; APPLICANT: Kerlavage, Anthony

; APPLICANT: Fraser, Claire M.

; APPLICANT: Kirkness, Ewen

; APPLICANT: Sutton, Granger

; APPLICANT: Gocayne, Jeannine

; APPLICANT: Liu, Li-Ing

; APPLICANT: Fitzgerald, Lisa

; APPLICANT: Adams, Mark

; APPLICANT: Lee, Norman

; APPLICANT: Fuldner, Rebecca

; APPLICANT: Fleischmann, Robert

; APPLICANT: Bult, Carol

; APPLICANT: Blake, Judy

; APPLICANT: White, Owen

; APPLICANT: Clayton, Rebecca

; APPLICANT: Pellegrino, Susan

; TITLE OF INVENTION: Human Genes, Sequences, and Expression

; NUMBER OF SEQUENCES: 16818

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart, &

; ADDRESSEE: Olstein

; STREET: 6 Becker Farm Road

; CITY: Roseland

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07068

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 5.0

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/196,363

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Olstein, Elliot

; REGISTRATION NUMBER: 24,025

; REFERENCE/DOCKET NUMBER: 325800-91

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (201) 994-1700

; TELEFAX: (201) 994-1744

; INFORMATION FOR SEQ ID NO: 1858:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 184 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

US-08-196-363-1858

Query Match 30.8%; Score 16; DB 5; Length 184;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GCATTGCCATATTGT 22

|||||

Db 165 GCATTGCCATATTGT 150

RESULT 6

US-08-196-363A-1858/c

; Sequence 1858, Application US/08196363A

; GENERAL INFORMATION:

; APPLICANT: Haseltine, William A.

; APPLICANT: Rosen, Craig A.

; APPLICANT: Ruben, Steven M.

; APPLICANT: Dillon, Patrick J.

; APPLICANT: Li, Haodong

; APPLICANT: Adams, Mark D.

; APPLICANT: Bult, Carol

; APPLICANT: Blake, Judith A.

; APPLICANT: Clayton, Rebecca

; APPLICANT: Fitzgerald, Lisa M.

; APPLICANT: Fleischmann, Robert

; APPLICANT: Fraser, Claire M.

; APPLICANT: Fuldner, Rebecca A.

; APPLICANT: Gocayne, Jeannine D.

; APPLICANT: Kerlavage, Anthony R.

; APPLICANT: Kirkness, Ewen F.

; APPLICANT: Lee, Norman H.

; APPLICANT: Liu, Li-Ing

; APPLICANT: Pellegrino, Susan M.

; APPLICANT: Sutton, Granger G.

; APPLICANT: White, Owen R.

; TITLE OF INVENTION: Human Genes, Sequences and Expression Products

; NUMBER OF SEQUENCES: 16820

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/196,363A

; FILING DATE: February 15, 1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Kenley K. Hoover

; REGISTRATION NUMBER: 40,302

; REFERENCE/DOCKET NUMBER: PO-06

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8439

; INFORMATION FOR SEQ ID NO: 1858:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 184 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

US-08-196-363A-1858

Query Match 30.8%; Score 16; DB 5; Length 184;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GCATTGCCATATTGT 22

|||||

Db 165 GCATTGCCATATTGT 150

RESULT 7

US-08-196-363-1858/c

; Sequence 1858, Application US/08196363D

; GENERAL INFORMATION:

; APPLICANT: Adams, Mark A.

; APPLICANT: Bult, Carol J.

; APPLICANT: Blake, Judith A.

; APPLICANT: Clayton, Rebecca

; APPLICANT: Fitzgerald, Lisa

; APPLICANT: Fleischmann, Robert D.
; APPLICANT: Fraser, Claire, M.
; APPLICANT: Fuldner, Rebecca A.
; APPLICANT: Gocayne, Jeannine D.
; APPLICANT: Kirlavage, Anthony R.
; APPLICANT: Kirkness, Even F.
; APPLICANT: Lee, Norman
; APPLICANT: Liu, Li-Ing
; APPLICANT: Pellegrino, Susan
; APPLICANT: Sutton III, Granger G.
; APPLICANT: White, Owen R.
; APPLICANT: Haseltine, William A.
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Li, Haodong
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products
; FILE REFERENCE: PO-06
; CURRENT APPLICATION NUMBER: US/08/196,363D
; CURRENT FILING DATE: 1994-02-15
; EARLIER APPLICATION NUMBER:
; EARLIER FILING DATE:
; NUMBER OF SEQ ID NOS: 16820
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1858
; LENGTH: 184
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (110)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (135)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145)
; OTHER INFORMATION: n equals a,t,g, or c
US-08-196-363-1858

Query Match 30.8%; Score 16; DB 5; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GCATTGCCATATTGT 22
|||||
Db 165 GCATTGCCATATTGT 150

RESULT 8
US-09-859-490-1858/c
; Sequence 1858, Application US/09859490
; GENERAL INFORMATION:
; APPLICANT: Adams, et al.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products
; FILE REFERENCE: PO-06C1
; CURRENT APPLICATION NUMBER: US/09/859,490
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 08/196,363
; PRIOR FILING DATE: 1994-02-15
; NUMBER OF SEQ ID NOS: 16820
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1858
; LENGTH: 184
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (110)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature

; LOCATION: (135)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (145)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-859-490-1858

Query Match 30.8%; Score 16; DB 32; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GCATTGCCATATTGT 22
|||||
Db 165 GCATTGCCATATTGT 150

RESULT 9
US-09-866-555-4699/c
; Sequence 4699, Application US/09866555
; GENERAL INFORMATION:
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Algate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Klee, Jennifer
; TITLE OF INVENTION: LUNG TUMOR-ASSOCIATED SEQUENCES AND METHODS OF USE
; FILE REFERENCE: 210121.580
; CURRENT APPLICATION NUMBER: US/09/866,555
; CURRENT FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 20487
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4699
; LENGTH: 184
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(184)
; OTHER INFORMATION: n = A,T,C or G
US-09-866-555-4699

Query Match 30.8%; Score 16; DB 33; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GCATTGCCATATTGT 22
|||||
Db 165 GCATTGCCATATTGT 150

RESULT 10
US-09-698-014-3804
; Sequence 3804, Application US/09698014
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Kingsbury, Gillian A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2014-001
; CURRENT APPLICATION NUMBER: US/09/698,014
; CURRENT FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: 60/162,363
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 6098
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3804
; LENGTH: 225
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-698-014-3804

```
Query Match 30.8%; Score 16; DB 27; Length 225;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 ATATTGTGAGGTCAC 30
|||||
Db 142 ATATTGTGAGGTCAC 157

RESULT 11
US-60-129-391-156/C
; Sequence 156, Application US/60129391
; GENERAL INFORMATION:
; APPLICANT: Keith et al.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR USE IN THE DIAGNOSIS AND TREATMENT
; FILE REFERENCE: 107196.138
; CURRENT APPLICATION NUMBER: US/60/129,391
; CURRENT FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 3674
; SEQ ID NO 156
; LENGTH: 319
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (49),(56),(57),(83),(246),(252),(266),(276)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
US-60-129-391-156

Query Match 30.8%; Score 16; DB 56; Length 319;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTCGCATTCGCAT 16
|||||
Db 47 CTCCTCGCATTCGCAT 32

RESULT 12
US-09-304-517A-269440/C
; Sequence 269440, Application US/09304517A
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)B
; CURRENT APPLICATION NUMBER: US/09/304,517A
; CURRENT FILING DATE: 1999-05-06
; NUMBER OF SEQ ID NOS: 295529
; SEQ ID NO 269440
; LENGTH: 328
; TYPE: DNA
; ORGANISM: Glycine max
US-09-304-517A-269440

Query Match 30.8%; Score 16; DB 17; Length 328;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GCATTGCCATATTGT 22
|||||
Db 279 GCATTGCCATATTGT 264

RESULT 13
US-09-371-146A-268282/C
; Sequence 268282, Application US/09371146A
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine
; TITLE OF INVENTION: ANNOTATED PLANT GENES
; FILE REFERENCE: 38-21(15097)C
```

```
; CURRENT APPLICATION NUMBER: US/09/371,146A
; CURRENT FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: US 09/304,517
; PRIOR FILING DATE: 1999-05-06
; NUMBER OF SEQ ID NOS: 294310
; SEQ ID NO 268282
; LENGTH: 328
; TYPE: DNA
; ORGANISM: Glycine max
US-09-371-146A-268282

Query Match 30.8%; Score 16; DB 17; Length 328;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GCATTGCCATATTGT 22
|||||
Db 279 GCATTGCCATATTGT 264

RESULT 14
US-09-565-240-18654/C
; Sequence 18654, Application US/09565240
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated wi
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15793)B
; CURRENT APPLICATION NUMBER: US/09/565,240
; CURRENT FILING DATE: 2000-05-03
; NUMBER OF SEQ ID NOS: 54005
; SEQ ID NO 18654
; LENGTH: 328
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3106-091-Q1-K1-D3
US-09-565-240-18654

Query Match 30.8%; Score 16; DB 22; Length 328;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GCATTGCCATATTGT 22
|||||
Db 279 GCATTGCCATATTGT 264

RESULT 15
US-09-985-678-269440/C
; Sequence 269440, Application US/09985678
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 16517.255/38-21(15097)F
; CURRENT APPLICATION NUMBER: US/09/985,678
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 09/304,517
; PRIOR FILING DATE: 1999-05-06
; NUMBER OF SEQ ID NOS: 295529
; SEQ ID NO 269440
; LENGTH: 328
; TYPE: DNA
; ORGANISM: Glycine max
US-09-985-678-269440

Query Match 30.8%; Score 16; DB 37; Length 328;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GCATTGCCATATTGT 22
|||||
```

Db 279 GCATTGCCATATTGT 264

Search completed: November 26, 2002, 20:08:04
Job time : 3786 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 15:36:20 ; Search time 31 Seconds
(without alignments)
646.000 Million cell updates/sec

Title: US-09-875-945-3
Perfect score: 52
Sequence: 1 ctctcgcattgccatattt.....gcadtagtatctgtgcaca 52

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 341543 seqs, 19255720 residues

Word size : 0

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_applications_NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	100.0	52	10	US-09-875-945-3
2	15	28.8	3924	12	US-10-044-090-188
3	15	28.8	198285	10	US-09-880-107-3814
4	14	26.9	226	10	US-09-878-574-5958
5	14	26.9	298	10	US-09-294-093B-4842
6	14	26.9	441	10	US-09-880-107-2934
7	14	26.9	463	10	US-09-920-300A-238
8	14	26.9	463	12	US-10-033-528-238
9	14	26.9	1033	10	US-09-770-445-186
10	14	26.9	1074	10	US-09-815-242-6743
11	14	26.9	2244	9	US-10-007-270-5
12	14	26.9	2887	9	US-10-007-270-3
13	14	26.9	3145	10	US-09-764-869-2190
14	14	26.9	3261	9	US-10-007-270-27
15	14	26.9	3330	9	US-10-007-270-1
16	14	26.9	3474	10	US-09-925-302-143
17	14	26.9	11821	10	US-09-764-877-2857
18	14	26.9	16518	9	US-09-764-868-1492
19	14	26.9	16798	10	US-09-070-927A-195

20	14	26.9	41907	10	US-09-967-013-5
c 21	14	26.9	66479	9	US-10-041-856-1
22	14	26.9	88191	10	US-09-799-799-3
c 23	13	25.0	17	10	US-09-866-108-7476
c 24	13	25.0	17	10	US-09-866-108-7477
c 25	13	25.0	17	10	US-09-866-108-7478
c 26	13	25.0	17	10	US-09-866-108-7479
c 27	13	25.0	17	10	US-09-866-108-7480
c 28	13	25.0	25	10	US-09-866-108-12368
c 29	13	25.0	25	10	US-09-866-108-12369
c 30	13	25.0	25	10	US-09-866-108-12370
c 31	13	25.0	25	10	US-09-866-108-12371
c 32	13	25.0	25	10	US-09-866-108-12372
c 33	13	25.0	25	10	US-09-866-108-12373
c 34	13	25.0	25	10	US-09-866-108-12374
c 35	13	25.0	25	10	US-09-866-108-12375
c 36	13	25.0	25	10	US-09-866-108-12376
c 37	13	25.0	25	10	US-09-866-108-12377
c 38	13	25.0	25	10	US-09-866-108-12378
c 39	13	25.0	25	10	US-09-866-108-12379
c 40	13	25.0	25	10	US-09-866-108-12380
c 41	13	25.0	105	10	US-09-864-761-33086
c 42	13	25.0	125	10	US-09-866-108-15691
c 43	13	25.0	162	10	US-09-988-598-2299
c 44	13	25.0	175	10	US-09-833-381-1823
c 45	13	25.0	197	10	US-09-864-761-30128

ALIGNMENTS

RESULT 1
US-09-875-945-3
; Sequence 3, Application US/09875945
; Patent No. US20020098169A1
; GENERAL INFORMATION:
; APPLICANT: METCON MEDICIN AB
; TITLE OF INVENTION: SMITH, Ulf
; FILE REFERENCE: 45513MH
; CURRENT APPLICATION NUMBER: US/09/875,945
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: SE 0002189-9
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: US 60/210,207
; PRIOR FILING DATE: 2000-06-08
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 3
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-875-945-3

Query Match	100.0%	Score 52;	DB 10;	Length 52;
Best Local Similarity	100.0%;	Pred. No. 1.4e-21;		
Matches 52;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	CTCCTCGCATTCGCATATTTGTGAGTGCTACTTGCAGTAGGTATCTGTGCACA	52	
DB	1	CTCCTCGCATTCGCATATTTGTGAGTGCTACTTGCAGTAGGTATCTGTGCACA	52	

RESULT 2
US-10-044-090-188
; Sequence 188, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09

```
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 188
; LENGTH: 3924
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 085713.1
; NAME/KEY: unsure
; LOCATION: 3902, 3912, 3918, 3920
; OTHER INFORMATION: a, t, c, g, or other
US-10-044-090-188

Query Match      28.8%; Score 15; DB 12; Length 3924;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 TTGCCATATTGTGA 24
Db 334 TTGCCATATTGTGA 348

RESULT 3
US-09-880-107-3814/c
; Sequence 3814, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3814
; LENGTH: 198285
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X87344
US-09-880-107-3814

Query Match      28.8%; Score 15; DB 10; Length 198285;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 GGTCACCTGCAGTAG 39
Db 75204 GGTCACCTGCAGTAG 75190

RESULT 4
US-09-878-574-5958/c
; Sequence 5958, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
```

```
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 5958
; LENGTH: 226
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701097566H1
US-09-878-574-5958

Query Match      26.9%; Score 14; DB 10; Length 226;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 CCATATTGTGAGG 26
Db 92 CCATATTGTGAGG 79

RESULT 5
US-09-294-093B-4842
; Sequence 4842, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 4842
; LENGTH: 298
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700355168H1
US-09-294-093B-4842

Query Match      26.9%; Score 14; DB 10; Length 298;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 CCATATTGTGAGG 26
Db 24 CCATATTGTGAGG 37

RESULT 6
US-09-880-107-2934/c
; Sequence 2934, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2934
; LENGTH: 441
```

;
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 R65593
; NAME/KEY: unsure
; LOCATION: (1)...(441)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-2934

Query Match 26.9%; Score 14; DB 10; Length 441;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CATTGCCATATTG 21
|||||
DB 65 CATTGCCATATTG 52

RESULT 7
US-09-920-300A-238/c
; Sequence 238, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 238
; LENGTH: 463
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-238

Query Match 26.9%; Score 14; DB 10; Length 463;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 GTGAGTCACTTGC 34
|||||
DB 153 GTGAGTCACTTGC 140

RESULT 8
US-10-033-528-238/c
; Sequence 238, Application US/10033528
; Patent No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033,528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 238
; LENGTH: 463
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-238

Query Match 26.9%; Score 14; DB 12; Length 463;
Best Local Similarity 100.0%; Pred. No. 29;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 GTGAGTCACTTGC 34
|||||
DB 153 GTGAGTCACTTGC 140

RESULT 9
US-09-770-445-186
; Sequence 186, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 186
; LENGTH: 1033
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1033)
; OTHER INFORMATION: n = A,T,C or G
US-09-770-445-186

Query Match 26.9%; Score 14; DB 10; Length 1033;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ATTGCCATATTGT 22
|||||
DB 959 ATTGCCATATTGT 972

RESULT 10
US-09-815-242-6743/c
; Sequence 6743, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in

;; TITLE OF INVENTION: Prokaryotes
;; FILE REFERENCE: ELITRA.011A
;; CURRENT APPLICATION NUMBER: US/09/815,242
;; CURRENT FILING DATE: 2001-03-21
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 6743
;; LENGTH: 1074
;; TYPE: DNA
;; ORGANISM: Enterococcus faecalis
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(1074)
US-09-815-242-6743

Query Match 26.9%; Score 14; DB 10; Length 1074;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 GTAGTATCTGTGC 49
Db 630 GTAGTATCTGTGC 617

RESULT 11
US-10-007-270-5/c
;; Sequence 5, Application US/10007270
;; Patent No. US20020160954A1
;; GENERAL INFORMATION:
;; APPLICANT: Hageman, Gregory S.
;; APPLICANT: Kuehn, Markus H.
;; APPLICANT: University of Iowa Research Foundation
;; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR OCULAR ABNORMALITIES
;; FILE REFERENCE: 020618-00012005
;; CURRENT APPLICATION NUMBER: US/10/007,270
;; CURRENT FILING DATE: 2001-11-08
;; PRIOR APPLICATION NUMBER: US 09/430,195
;; PRIOR FILING DATE: 1999-10-29
;; PRIOR APPLICATION NUMBER: US 09/183,972
;; PRIOR FILING DATE: 1998-10-29
;; NUMBER OF SEQ ID NOS: 37
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 5
;; LENGTH: 2244
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: Human IPM 150 cDNA sequence, isoform C
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)..(2244)
;; OTHER INFORMATION: n is a, c, g, or t.
US-10-007-270-5

Query Match 26.9%; Score 14; DB 9; Length 2244;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAGGTCACTTGCAG 36

Db 1178 GAGGTCACTTGCAG 1165

RESULT 12
US-10-007-270-3/c
;; Sequence 3, Application US/10007270
;; Patent No. US20020160954A1
;; GENERAL INFORMATION:
;; APPLICANT: Hageman, Gregory S.
;; APPLICANT: Kuehn, Markus H.
;; APPLICANT: University of Iowa Research Foundation
;; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR OCULAR ABNORMALITIES
;; FILE REFERENCE: 020618-00012005
;; CURRENT APPLICATION NUMBER: US/10/007,270
;; CURRENT FILING DATE: 2001-11-08
;; PRIOR APPLICATION NUMBER: US 09/430,195
;; PRIOR FILING DATE: 1999-10-29
;; PRIOR APPLICATION NUMBER: US 09/183,972
;; PRIOR FILING DATE: 1998-10-29
;; NUMBER OF SEQ ID NOS: 37
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 3
;; LENGTH: 2887
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: Human IPM 150 cDNA sequence, isoform B
US-10-007-270-3

Query Match 26.9%; Score 14; DB 9; Length 2887;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAGGTCACTTGCAG 36
Db 715 GAGGTCACTTGCAG 702

RESULT 13
US-09-764-869-2190/c
;; Sequence 2190, Application US/09764869
;; Patent No. US20020061521A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: PC007
;; CURRENT APPLICATION NUMBER: US/09/764,869
;; CURRENT FILING DATE: 2001-01-17
;; PRIOR APPLICATION data removed - refer to PALM or file wrapper
;; NUMBER OF SEQ ID NOS: 2442
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 2190
;; LENGTH: 3145
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-764-869-2190

Query Match 26.9%; Score 14; DB 10; Length 3145;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 TTGTGAGGTCACTT 32
Db 600 TTGTGAGGTCACTT 587

RESULT 14
US-10-007-270-27/c
;; Sequence 27, Application US/10007270
;; Patent No. US20020160954A1
;; GENERAL INFORMATION:
;; APPLICANT: Hageman, Gregory S.

; APPLICANT: Kuehn, Markus H.
; APPLICANT: University of Iowa Research Foundation
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR OCULAR ABNORMALITIES
; FILE REFERENCE: 020618-000120US
; CURRENT APPLICATION NUMBER: US/10/007,270
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 09/430,195
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 09/183,972
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 3261
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Human IPM 150 isoform A variant cDNA sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (128)..(2440)
US-10-007-270-27

Query Match 26.9%; Score 14; DB 9; Length 3261;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAGGTCACCTGCAG 36
|||||
DB 1093 GAGGTCACCTGCAG 1080

RESULT 15
US-10-007-270-1/c
; Sequence 1, Application US/10007270
; Patent No. US20020160954A1
; GENERAL INFORMATION:
; APPLICANT: Hageman, Gregory S.
; APPLICANT: Kuehn, Markus H.
; APPLICANT: University of Iowa Research Foundation
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR OCULAR ABNORMALITIES
; FILE REFERENCE: 020618-000120US
; CURRENT APPLICATION NUMBER: US/10/007,270
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 09/430,195
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 09/183,972
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3330
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Human IPM 150 cDNA, isoform A
US-10-007-270-1

Query Match 26.9%; Score 14; DB 9; Length 3330;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAGGTCACCTGCAG 36
|||||
DB 1158 GAGGTCACCTGCAG 1145

Search completed: November 26, 2002, 20:13:49
Job time : 69 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 14:32:35 ; Search time 36 seconds
(without alignments)
442.978 Million cell updates/sec

Title: US-09-875-945-3
Perfect score: 52
Sequence: 1 cctctgcattgcattt.....gcagtaggtatctgtgcaca 52

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	30.8	9244	US-08-961-527-68	Sequence 68, Appl
2	14	26.9	344	US-08-961-527-352	Sequence 352, App
3	14	26.9	1286	US-09-149-476-242	Sequence 242, App
4	14	26.9	1374	US-08-123-761A-2	Sequence 2, Appli
5	14	26.9	1491	5486473-3	Patent No. 5486473
6	14	26.9	1493	US-08-752-307B-6	Sequence 6, Appli
7	14	26.9	1493	US-09-707-802-6	Sequence 6, Appli
8	14	26.9	1493	US-09-991-326-6	Sequence 6, Appli
9	14	26.9	1628	US-09-147-522-3	Sequence 3, Appli
10	14	26.9	2403	US-08-700-013B-26	Sequence 26, Appli
11	14	26.9	2817	US-08-655-836-4	Sequence 4, Appli
12	14	26.9	2817	US-09-020-753-4	Sequence 4, Appli
13	14	26.9	4177	US-09-023-082A-23	Sequence 23, Appli
14	14	26.9	5000	US-09-147-522-5	Sequence 5, Appli
15	14	26.9	5198	US-08-123-761A-1	Sequence 1, Appli
16	14	26.9	6735	US-08-961-527-104	Sequence 104, App
17	14	26.9	11298	US-07-869-933-31	Sequence 31, Appli
18	14	26.9	11298	US-07-869-933-31	Sequence 31, Appli
19	14	26.9	11298	US-08-201-879A-2	Sequence 2, Appli
20	14	26.9	11298	US-08-201-879A-2	Sequence 2, Appli
21	14	26.9	11298	US-09-103-663-31	Sequence 31, Appli
22	14	26.9	11298	US-09-103-663-31	Sequence 31, Appli
23	14	26.9	11443	US-08-961-527-49	Sequence 49, Appli
24	14	26.9	45546	US-09-146-053-6	Sequence 6, Appli
25	14	26.9	59065	US-09-813-817-3	Sequence 3, Appli
26	14	26.9	59065	US-09-978-197-3	Sequence 3, Appli
27	13	25.0	28	US-08-859-998-907	Sequence 907, App

c	28	13	25.0	28	4	US-09-225-928-907	Sequence 907, App
	29	13	25.0	41	4	US-09-351-814-54	Sequence 54, Appl
	30	13	25.0	234	4	US-08-905-223-164	Sequence 164, App
	31	13	25.0	423	1	US-08-470-179-159	Sequence 159, App
	32	13	25.0	423	1	US-08-470-179-177	Sequence 177, App
	33	13	25.0	427	1	US-08-470-179-195	Sequence 195, App
c	34	13	25.0	437	3	US-09-115-061-20	Sequence 20, Appl
c	35	13	25.0	437	4	US-09-261-079-20	Sequence 20, Appl
c	36	13	25.0	451	4	US-08-998-416-196	Sequence 196, App
c	37	13	25.0	514	3	US-08-875-573-5	Sequence 5, Appli
- c	38	13	25.0	521	4	US-09-404-879A-34	Sequence 34, Appl
	39	13	25.0	548	4	US-09-484-970B-102	Sequence 102, App
	40	13	25.0	620	2	US-08-215-089-1	Sequence 1, Appli
	41	13	25.0	620	5	PCT-US95-03384-1	Sequence 1, Appli
c	42	13	25.0	975	4	US-09-381-488-6	Sequence 6, Appli
c	43	13	25.0	993	4	US-09-134-001C-1165	Sequence 1165, Ap
c	44	13	25.0	1024	3	US-08-934-494-3	Sequence 3, Appli
c	45	13	25.0	1024	3	US-09-143-068-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-961-527-68
; Sequence 68, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9244 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-68

Query Match 30.8%; Score 16; DB 4; Length 9244;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 CTTGCAGTAGGTATCT 45
|||||

DB 755 CTTGCAGTAGGTATCT 770

```
RESULT 2
US-08-961-527-352/c
; Sequence 352, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 352:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-352
Query Match 26.98; Score 14; DB 4; Length 344;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ATTGCCATATTGT 22
Db 58 ATTGCCATATTGT 45

RESULT 3
US-09-149-476-242/c
; Sequence 242, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
```

```
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
```

; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,631
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,845
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,892
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,761
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/047,595
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,599
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,588
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,585
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,586
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,590
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,594
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,589
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,593
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,614
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,578
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,576

; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/047,501
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,670
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/056,632
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,664
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,876
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,881
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,909
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,875
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,862
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,887
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,908
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/057,650
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/056,884
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,669
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/049,610
; EARLIER FILING DATE: 1997-06-13
; EARLIER APPLICATION NUMBER: 60/061,060
; EARLIER FILING DATE: 1997-10-02

Query Match 26.9%; Score 14; DB 4; Length 1286;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TCGCATGCCATAT 18
|||||
DB 485 TCGCATGCCATAT 472

RESULT 4
US-08-123-761A-2/c
; Sequence 2, Application US/08123761A
; Patent No. 5589611
; GENERAL INFORMATION:
; APPLICANT: Briggs, Steven P.
; APPLICANT: Johal, Gurmukh S.
; TITLE OF INVENTION: A DISEASE RESISTANCE GENE FROM MAIZE AS
; TITLE OF INVENTION: A SELECTABLE MARKER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
; STREET: 700 Capital Square, 400 Locust Street
; CITY: Des Moines
; STATE: Iowa
; COUNTRY: USA
; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/123,761A
; FILING DATE: 17-SEP-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Roth, Michael J.

REGISTRATION NUMBER: 29,342
REFERENCE/DOCKET NUMBER: 212-US
TELEPHONE: (515) 248-4800
TELEFAX: (515) 248-4844
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1374 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-123-761A-2

Query Match 26.9%; Score 14; DB 1; Length 1374;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCCTCCGATTGCC 14
|||||

Db 402 CTCCTCGCATTGCC 389

RESULT 5
5486473-3
; Patent No. 5486473
; APPLICANT: FUKITA, HIROYUKI; YOSHIDA, IWAO; TAKAGI, MITSUO;
; MANABE, SADAQ; FUKAI, KONOSUKE
; TITLE OF INVENTION: A DNA CODING FOR A FLAVIVIRUS ANTIGEN
; NUMBER OF SEQUENCES: 16
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,049
; FILING DATE: 09-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 809,255
; FILING DATE: 18-DEC-1991
; APPLICATION NUMBER: 279,685
; FILING DATE: 05-DEC-1988
; APPLICATION NUMBER: 932,419
; FILING DATE: 19-NOV-1986
; SEQ ID NO: 3;
; LENGTH: 1491
5486473-3

Query Match 26.9%; Score 14; DB 6; Length 1491;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 TTGCCATATTTGTG 23
|||||

Db 416 TTGCCATATTTGTG 429

RESULT 6
US-08-752-307B-6
; Sequence 6, Application US/08752307B
; Patent No. 5952171
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Gearing, David P.
; APPLICANT: Levinson, Douglas A.
; TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES
; TITLE OF INVENTION: ENCODING NOVEL SECRETED OR MEMBRANE-ASSOCIATED PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,307B
FILING DATE: 19-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Meiklejohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 09404/020001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1493 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 99...1493
US-08-752-307B-6

Query Match 26.9%; Score 14; DB 2; Length 1493;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 ATTGTGAGGTGCAC 30
|||||

Db 975 ATTGTGAGGTGCAC 988

RESULT 7
US-09-707-802-6
; Sequence 6, Application US/09707802
; Patent No. 6391586
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; Gearing, David P.
; Levinson, Douglas A.
; TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES
; TITLE OF INVENTION: ENCODING NOVEL SECRETED OR MEMBRANE-ASSOCIATED PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/707,802
; FILING DATE: 07-No. 6391586-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/752,307
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 09404/020001

```
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 99...1493
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-707-802-6

Query Match 26.9%; Score 14; DB 4; Length 1493;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 ATTGTGAGGTCAC 30
Db 975 ATTGTGAGGTCAC 988
|||||

RESULT 8
US-09-991-326-6
; Sequence 6, Application US/09991326
; Patent No. 6395872
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; Gearing, David P.
; Levinson, Douglas A.
; TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/991.326
; FILING DATE: 21-Nov. 6395872-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/752.307
; FILING DATE: 19-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 09404/020002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 99...1493
```

```
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-991-326-6

Query Match 26.9%; Score 14; DB 4; Length 1493;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 ATTGTGAGGTCAC 30
Db 975 ATTGTGAGGTCAC 988
|||||

RESULT 9
US-09-147-522-3
; Sequence 3, Application US/09147522
; Patent No. 6107069
; GENERAL INFORMATION:
; APPLICANT: MAGAGNIN, SIMONA
; APPLICANT: BENATTI, LUCA
; APPLICANT: CINL, MASSIMO
; APPLICANT: SPECIALE, CARMELA
; APPLICANT: COVINI, NEVIE
; TITLE OF INVENTION: RECOMBINANT KYNURENINE-3-HYDROXYLASE ENZYME AND
; FILE REFERENCE: 0769-0408-0PCT
; CURRENT APPLICATION NUMBER: US/09/147.522
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: PCT/EP7/03589
; EARLIER FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1628
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (34)..(1494)
US-09-147-522-3

Query Match 26.9%; Score 14; DB 3; Length 1628;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 CATTGCCATATTG 21
Db 1564 CATTGCCATATTG 1577
|||||

RESULT 10
US-08-700-013B-26
; Sequence 26, Application US/08700013B
; Patent No. 5919653
; GENERAL INFORMATION:
; APPLICANT: Albert, Vivian R.
; APPLICANT: Kowalski, Leslie R.Z.
; APPLICANT: Borden, Laurence A.
; APPLICANT: McKelvy, Jeffrey F.
; TITLE OF INVENTION: Human Glycine Transporter
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/700,013B
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bloom, Allen
REGISTRATION NUMBER: 29,135
REFERENCE/DOCKET NUMBER: 317743-108
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-520-3214
TELEFAX: 609-520-3259
TELEX:
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 2403 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-700-013B-26

Query Match 26.9%; Score 14; DB 2; Length 2403;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CATTGCCATATTG 21
|||||

DB 1935 CATTGCCATATTG 1948

RESULT 11
US-08-655-836-4
; Sequence 4, Application US/08655836
; Patent No. 582486
; GENERAL INFORMATION:
; APPLICANT: Borden, Laurence A.
; APPLICANT: De Vivo, Michael
; APPLICANT: Yokoyama, Midori
; APPLICANT: Albert, Vivian R.
; TITLE OF INVENTION: GLYCINE TRANSPORTER-TRANSFECTED CELLS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: P.O. Box 5218
; CITY: Princeton
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patenting Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/655,836
; FILING DATE: 31-MAY-1996
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29135
; REFERENCE/DOCKET NUMBER: 317743-108
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2817 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna

US-08-655-836-4
Query Match 26.9%; Score 14; DB 1; Length 2817;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CATTGCCATATTG 21
|||||

DB 2142 CATTGCCATATTG 2155

RESULT 12
US-09-020-753-4
; Sequence 4, Application US/09020753
; Patent No. 5968823
; GENERAL INFORMATION:
; APPLICANT: Borden, Laurence A.
; APPLICANT: Devivo, Michael
; APPLICANT: Yokoyama, Midori
; APPLICANT: Albert, Vivian R.
; TITLE OF INVENTION: Glycine Transporter-Transfected
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,753
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: 08/655,836
; APPLICATION NUMBER: 31-MAY-1996
; FILING DATE: 31-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2817 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-020-753-4

Query Match 26.9%; Score 14; DB 2; Length 2817;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CATTGCCATATTG 21
|||||

DB 2142 CATTGCCATATTG 2155

RESULT 13
US-09-023-082A-23/c
; Sequence 23, Application US/09023082A
; Patent No. 6077692
; GENERAL INFORMATION:
; APPLICANT: RUBEN, STEVEN M.

APPLICANT: JIMENEZ, PABLO
APPLICANT: DUAN, D. ROXANNE
APPLICANT: RAMPY, MARK A.
APPLICANT: MENDRICK, DONNA
APPLICANT: ZHANG, JUN
APPLICANT: NI, JIAN
APPLICANT: MOORE, PAUL A.
APPLICANT: COLEMAN, TIMOTHY A.
APPLICANT: GRUBER, JOACHIM R.
APPLICANT: DILLON, PATRICK J.
APPLICANT: GENTZ, REINER L.
TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVE, NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,082A
FILING DATE: 13-FEB-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/01790
FILING DATE: 14-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/461,195
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/023,852
FILING DATE: 13-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/039,045
FILING DATE: 28-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/862,432
FILING DATE: 23-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/910,875
FILING DATE: 13-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/055,561
FILING DATE: 13-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: STEFFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0360008/EKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 4177 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: both
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 593..1216
US-09-023-082A-23

Query Match 26.9%; Score 14; DB 3; Length 4177;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GCATTGCCATATT 20
|||||
Db 2096 GCATTGCCATATT 2083
RESULT 14
US-09-147-522-5
; Sequence 5, Application US/09147522
; Patent No. 6107069
; GENERAL INFORMATION:
; APPLICANT: MAGAGNIN, SIMONA
; APPLICANT: BENATTI, LUCA
; APPLICANT: CINI, MASSIMO
; APPLICANT: SPECIALE, CARMELA
; APPLICANT: COVINI, NEVIE
; TITLE OF INVENTION: RECOMBINANT KYNURENINE-3-HYDROXYLASE ENZYME AND
; TITLE OF INVENTION: PROCESS FOR ITS PREPARATION
; FILE REFERENCE: 0769-0408-OPCT
; CURRENT APPLICATION NUMBER: US/09/147,522
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: PCT/EP7/03589
; EARLIER FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 5000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (47)..(1507)
US-09-147-522-5
Query Match 26.9%; Score 14; DB 3; Length 5000;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 CATTGCCATATTG 21
|||||
Db 1577 CATTGCCATATTG 1590
RESULT 15
US-08-123-761A-1/c
; Sequence 1, Application US/08123761A
; Patent No. 5589611
; GENERAL INFORMATION:
; APPLICANT: Briggs, Steven P.
; APPLICANT: Johal, Gurmukh S.
; TITLE OF INVENTION: A DISEASE RESISTANCE GENE FROM MAIZE AS
; TITLE OF INVENTION: A SELECTABLE MARKER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
; STREET: 700 Capital Square, 400 Locust Street
; CITY: Des Moines
; STATE: Iowa
; COUNTRY: USA
; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/123,761A
; FILING DATE: 17-SEP-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Roth, Michael J.
; REGISTRATION NUMBER: 29,342
; REFERENCE/DOCKET NUMBER: 212-US
; TELECOMMUNICATION INFORMATION:

```
; TELEPHONE: (515) 248-4800
; TELEFAX: (515) 248-4844
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5198 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-123-761A-1

Query Match      26.9%; Score 14; DB 1; Length 5198;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCCTCGCATTGCC 14
   |||||
Db 3774 CTCCTCGCATTGCC 3761
```

Search completed: November 26, 2002, 19:04:48
Job time : 40 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 12:46:30 ; Search time 2198 Seconds
(without alignments)
383.151 Million cell updates/sec

Title: US-09-875-945-3
Perfect score: 52
Sequence: 1 ctctctgcattgccatattt.....gcagtaggtatctgtgcaca 52

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST.*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estmu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_hic.*
- 9: gb_est1.*
- 10: gb_est2.*
- 11: gb_hic.*
- 12: gb_est3.*
- 13: gb_est4.*
- 14: gb_est5.*
- 15: em_estfun.*
- 16: em_estom.*
- 17: gb_gss.*
- 18: em_gss_hum.*
- 19: em_gss_inv.*
- 20: em_gss_pln.*
- 21: em_gss_vrt.*
- 22: em_gss_fun.*
- 23: em_gss_mam.*
- 24: em_gss_mus.*
- 25: em_gss_other.*
- 26: em_gss_pro.*
- 27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
c 1	20	38.5	468	17	AZ098084	AZ098084 RPCI-23-4
c 2	20	38.5	558	17	AZ112424	AZ112424 RPCI-23-4
c 3	19	36.5	660	17	AZ823698	AZ823698 2M0097G22
c 4	18	34.6	234	17	AZ802259	AZ802259 2M0061B01
c 5	18	34.6	466	17	AZ372930	AZ372930 1M0125E03
c 6	18	34.6	511	10	AV990836	AV990836 AV990836

c 7	18	34.6	554	14	BQ630245	BQ630245 sap99g05.
c 8	18	34.6	556	13	BM528831	BM528831 sak69a04.
c 9	18	34.6	571	10	AV678795	AV678795 AV678795
10	18	34.6	581	10	AV981801	AV981801 AV981801
11	18	34.6	609	17	BH428985	BH428985 BOGL182TR
c 12	18	34.6	614	12	BF519194	BF519194 EST456655
c 13	18	34.6	634	14	BQ141233	BQ141233 NF017D09P
c 14	18	34.6	645	10	AW348152	AW348152 GM210001A
c 15	18	34.6	875	17	BH680500	BH680500 BOMG22TF
c 16	17	32.7	241	10	BB198743	BB198743 BB198743
c 17	17	32.7	419	17	AQ148299	AQ148299 HS_3126_A
c 18	17	32.7	430	12	BF523688	BF523688 UI-R-Y0-a
c 19	17	32.7	442	10	BB850846	BB850846 BB850846
c 20	17	32.7	454	13	BI470963	BI470963 sah87g06.
c 21	17	32.7	605	10	BE375317	BE375317 601225580
c 22	17	32.7	629	10	BB659929	BB659929 BB659929
c 23	17	32.7	641	10	BB612418	BB612418 BB612418
c 24	17	32.7	699	10	BB627445	BB627445 BB627445
c 25	17	32.7	702	12	BE916163	BE916163 601668784
c 26	17	32.7	739	13	BI100381	BI100381 602885857
c 27	17	32.7	2745	11	AK019482	AK019482 Mus muscu
c 28	17	32.7	3204	11	AK004640	AK004640 Mus muscu
c 29	16	30.8	44	10	AV838294	AV838294 AV838294
c 30	16	30.8	149	17	BH192522	BH192522 TC3-7C9.T
c 31	16	30.8	159	13	BM002290	BM002290 1031102B1
c 32	16	30.8	161	17	BH193894	BH193894 TC3-4G11.
c 33	16	30.8	167	17	BH193898	BH193898 TC3-4G11.
c 34	16	30.8	184	9	AA326897	AA326897 EST30128
c 35	16	30.8	279	14	BP027196	BP027196 BF027196
c 36	16	30.8	309	13	BI542962	BI542962 949023C02
c 37	16	30.8	322	13	BI595811	BI595811 949023C02
c 38	16	30.8	341	13	BI096411	BI096411 949016C11
c 39	16	30.8	374	17	AZ867860	AZ867860 2M0178C24
c 40	16	30.8	386	13	BI998426	BI998426 1031057D0
c 41	16	30.8	394	10	AV838274	AV838274 AV838274
c 42	16	30.8	438	17	BH617845	BH617845 SALK_0379
c 43	16	30.8	442	17	AZ144201	AZ144201 SP_0020_B
c 44	16	30.8	447	13	BM268213	BM268213 MEST378-C
c 45	16	30.8	459	13	BI097623	BI097623 949016C11

ALIGNMENTS

RESULT 1
AZ098084/c
LOCUS
DEFINITION
DNA sequence.
ACCESSION
AZ098084
VERSION
AZ098084.1 GI:7751140
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 468)
AUTHORS
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
,B., Levins,W., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
TITLE
Mouse BAC End Sequences from Library RPCI-23
JOURNAL
Unpublished (1999)
COMMENT
Other GSSs: RPCI-23-462A6.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@jgong.med.buffalo.edu). Clones may be purchased from

AZ098084
RPCI-23-462A6.TV
RPCI-23
Mus musculus
genomic clone
RPCI-23-462A6,
468 bp
DNA
linear
GSS 09-MAY-2000

BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
 or from Resea ch Genetics ([info@resgen.com](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 462 row: A column: 6
 Seq primer: T7
 Class: BAC ends.

Location/Qualifiers

1. .468
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-462A6"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
 EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBACE3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 155 a 85 c 108 g 120 t
 ORIGIN

Query Match 38.5%; Score 20; DB 17; Length 468;
 Best Local Similarity 100.0%; Pred. No. 2.3;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 CCATATTGTGAGGTCACTT 32
 |||||
 Db 358 CCATATTGTGAGGTCACTT 339

RESULT 2

AZ112424/c
 LOCUS
 DEFINITION AZ112424 558 bp DNA linear GSS 12-MAY-2000
 Mus musculus genomic clone RPCI-23-462C8,
 DNA sequence.

ACCESSION AZ112424
 VERSION AZ112424.1 GI:7771310
 KEYWORDS GSS.
 SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 558)
 Zhao,S., Niernan,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
 B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
 and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Other_GSSs: RPCI-23-462C8.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
 or from Resea ch Genetics ([info@resgen.com](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 462 row: C column: 8
 Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1. .558
 /organism="Mus musculus"
 /strain="C57BL/6J"

FEATURES

source

/db_xref="taxon:10090"
 /clone="RPCI-23-462C8"
 /clone_lib="RPCI-23"
 /sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
 EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBACE3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 184 a 97 c 134 g 143 t
 ORIGIN

Query Match 38.5%; Score 20; DB 17; Length 558;
 Best Local Similarity 100.0%; Pred. No. 2.5;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 CCATATTGTGAGGTCACTT 32
 |||||
 Db 355 CCATATTGTGAGGTCACTT 336

RESULT 3

AZ823698/c

LOCUS
 DEFINITION AZ823698 660 bp DNA linear GSS 20-FEB-2001
 Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC2M0097G22 R, DNA sequence.

ACCESSION AZ823698

VERSION AZ823698.1 GI:12993606

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 660)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0097 row: G column: 22

Seq primer: CACACAGGAACACGCTATGACC

Class: plasmid ends

High quality sequence stop: 660.

Location/Qualifiers

1. .660

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC2M0097G22"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource

(<http://www.jax.org/resources/documents/dnares/>). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 175 a 147 c 127 g 211 t
ORIGIN

Query Match 36.5%; Score 19; DB 17; Length 660;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 CAGTAGGTATCTGTGCACA 52
|||||
Db 247 CAGTAGGTATCTGTGCACA 229

RESULT 4

AZ802259/c
LOCUS 234 bp DNA linear GSS 16-FEB-2001
DEFINITION 2M0061B01F Mouse 10kb plasmid UUGCLM library Mus musculus genomic clone UUGC2M0061B01 F, DNA sequence.

ACCESSION AZ802259
VERSION AZ802259.1 GI:12954498
KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 234)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

Unpublished (2000)

JOURNAL

COMMENT

Contact: Robert B. Weiss
University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0061 row: B column: 01

Seq primer: CGTTGTAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 234.

Location/Qualifiers

FEATURES

source

1..234
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0061B01"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"

/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 72 a 46 c 33 g 83 t
ORIGIN

Query Match 34.6%; Score 18; DB 17; Length 234;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 ATATTTGTGAGGTCACTT 32
|||||
Db 112 ATATTTGTGAGGTCACTT 95

RESULT 5

AZ372930/c
LOCUS 466 bp DNA linear GSS 02-OCT-2000
DEFINITION 1M0125E03F Mouse 10kb plasmid UUGCLM library Mus musculus genomic clone UUGCLM0125E03 F, DNA sequence.

ACCESSION AZ372930
VERSION AZ372930.1 GI:10486630
KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 466)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

Unpublished (2000)

JOURNAL

COMMENT

Contact: Robert B. Weiss
University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0125 row: E column: 03

Seq primer: CGTTGTAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 466.

Location/Qualifiers

FEATURES

source

1..466
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0125E03"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"

/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pW42 (gll4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 144 a 127 c 92 g 103 t
ORIGIN
Query Match 34.6%; Score 18; DB 17; Length 466;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 ATTGCCATATTGTGAGG 26
|||||
Db 287 ATTGCCATATTGTGAGG 270

RESULT 6
AV990836 511 bp mRNA linear EST 14-MAR-2002
LOCUS
DEFINITION AV990836 Nori Satoh unpublished cDNA library, larva Ciona
Intestinalis cDNA clone cllv23n10 5', mRNA sequence.

ACCESSION AV990836
VERSION AV990836.1 GI:19479377
KEYWORDS EST.
SOURCE Ciona intestinalis.
ORGANISM Ciona intestinalis.

REFERENCE 1 (bases 1 to 511)
AUTHORS Satoh.N., Satou.Y., Kohara.Y. and Shin-I.T.
TITLE Expressed genes in Ciona intestinalis
JOURNAL Unpublished (2000)
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES
source
1..511
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="cillv23n10"
/clone_lib="Nori Satoh unpublished cDNA library, larva"
/tissue_type="whole animal"
/dev_stage="larva"
/note="Vector: pBluescript SK"
BASE COUNT 166 a 102 c 111 g 132 t
ORIGIN

Query Match 34.6%; Score 18; DB 10; Length 511;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 TCGCATTGCCATATTGT 22
|||||
Db 459 TCGCATTGCCATATTGT 476

RESULT 7
B0630245/c 554 bp mRNA linear EST 02-JUL-2002
LOCUS
DEFINITION B0630245 Glycine max cDNA clone SOYBEAN CLONE ID: 5',
sap99g05.yl Gm-cl045 similar to TR:O9XIS3 Q9XIS3 LECTIN-LIKE PROTEIN KINASE. ;, mRNA

sequence.
B0630245 GI:21677894
VERSION
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

REFERENCE 1 (bases 1 to 554)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpeliding,J., Coryell,V., Khanna
A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
R., Waterston,R. and Wilson,R.
Public Soybean EST Project

TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntville, AL 35801 For further information
call: (800)-533-4363 or contact: ccu@resgen.com web site:
www.resgen.com

Seq primer: -40RP from Gibco
High quality sequence stop: 430.
Location/Qualifiers
1..554
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID:"
/clone_lib="Gm-cl045"
/tissue_type="Hypocotyl, 9-10 day old etiolated seedlings"
/lab_host="DH10B"

/note="Vector: pBluescriptII SK+; Site_1: EcoRI; Site_2:
XhoI; This cDNA library was constructed from mRNA isolated
from etiolated hypocotyl tissue of 9-10 day old seedlings
of the cultivar Williams 82. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) primer with a XhoI restriction site. EcoRI
adapters were ligated to the blunt-ended cDNA fragments
followed by digestion with EcoRI and XhoI. The cDNA
fragments were directionally cloned into the EcoRI-XhoI
restriction site of the pBluescript vector. The ligated
cDNA fragments were transformed into DH10B host cells
(Gibco BRL). This library was constructed by Dr. Randy
Shoemaker."

BASE COUNT 155 a 115 c 91 g 193 t
ORIGIN

Query Match 34.6%; Score 18; DB 14; Length 554;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 TTGTGAGGTCACTTGCG 36
|||||
Db 189 TTGTGAGGTCACTTGCG 172

RESULT 8
BM528831/c 556 bp mRNA linear EST 19-FEB-2002
LOCUS
DEFINITION BM528831 sak69a04.yl Gm-cl036 Glycine max cDNA clone SOYBEAN CLONE ID:
Gm-cl036-8095 5' similar to TR:O9XIS3 Q9XIS3 LECTIN-LIKE PROTEIN
KINASE. ;, mRNA sequence.
ACCESSION BM528831
VERSION BM528831.1 GI:18735559

```

KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max

REFERENCE 1 (bases 1 to 556)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact: ccu@resgen.com web site:
www.resgen.com
Seq primer: -40RP from Gibco
High quality sequence stop: 394.
FEATURES source
Location/Qualifiers
1..556
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl036-8095"
/clone_lib="Gm-cl036"
/tissue_type="somatic embryos cultured on MSD 20"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This
cDNA library was constructed from mRNA isolated from
somatic embryos (age ranging from 2 months to 9 months)
cultured on MSD 20. The library was prepared using the
Life Technologies pSuperScript cDNA library construction
kit. Complementary DNA was synthesized from mRNA using a
poly (dT) sequence with a NotI restriction site. SalI
linkers adapters were ligated to the blunt-ended cDNA
fragments followed by NotI digestion. The cDNA fragments
were directionally cloned into the NotI-SalI restriction
site of the pSPORT1 vector. The ligated cDNA fragments
were transformed into E.coli ElectroMax DH10B host cells.
This library was constructed in the laboratory of Dr. Lila
Vodkin by Anu Khanna at the University of Illinois at
Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"
BASE COUNT 142 a 116 c 169 g 129 t
ORIGIN
Query Match 34.6%; Score 18; DB 13; Length 556;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 TTGTGAGGTCACCTGCAG 36
|||||
Db 542 TTGTGAGGTCACCTGCAG 525

RESULT 9
AV678795/c
LOCUS AV678795
DEFINITION AV678795 Nori Satoh unpublished cDNA library Ciona intestinalis
cDNA clone rcitb11k10 3', mRNA sequence.
ACCESSION AV678795
VERSION AV678795.1 GI:10116794
KEYWORDS EST.
SOURCE Ciona intestinalis.

ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cionidae; Ciona.
REFERENCE 1 (bases 1 to 571)
AUTHORS Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
TITLE Expressed genes in Ciona intestinalis
JOURNAL Unpublished (2000)
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES source
Location/Qualifiers
1..571
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="rcitb11k10"
/clone_lib="Nori Satoh unpublished cDNA library"
/tissue_type="whole animal"
/dev_stage="tailbud"
/note="Vector: pBluescript SK"
BASE COUNT 154 a 120 c 116 g 181 t
ORIGIN
Query Match 34.6%; Score 18; DB 10; Length 571;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TCGCATTGCCATATTGT 22
|||||
Db 27 TCGCATTGCCATATTGT 10

RESULT 10
AV981801
LOCUS AV981801
DEFINITION AV981801 Nori Satoh unpublished cDNA library, larva Ciona
intestinalis cDNA clone cilv36b06 5', mRNA sequence.
ACCESSION AV981801
VERSION AV981801.1 GI:19471200
KEYWORDS EST.
SOURCE Ciona intestinalis.
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cionidae; Ciona.
REFERENCE 1 (bases 1 to 581)
AUTHORS Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
TITLE Expressed genes in Ciona intestinalis
JOURNAL Unpublished (2000)
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES source
Location/Qualifiers
1..581
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="cilv36b06"
/clone_lib="Nori Satoh unpublished cDNA library, larva"
/tissue_type="whole animal"
/dev_stage="larva"
/note="Vector: pBluescript SK"
BASE COUNT 193 a 113 c 121 g 153 t
ORIGIN
Query Match 34.6%; Score 18; DB 10; Length 581;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      5  TCGCATTGCCATATTGCT 22
          |||
Db      529 TCGCATTGCCATATTGCT 546

RESULT 11
BH428985
LOCUS   BH428985
DEFINITION BOGLI82TR BOGL Brassica oleracea genomic clone BOGLI82, DNA
sequence.
ACCESSION BH428985
VERSION   BH428985.1 GI:17614706
KEYWORDS  GSS.
SOURCE    Brassica oleracea.
ORGANISM  Brassica oleracea
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 609)
          Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
          Whole genome shotgun sequencing of Brassica oleracea
          unpublished (2001)
          Other_GSSs: BOGLI82TF
          Contact: Chris Town
          TIGR
          9712 Medical Center Drive, Rockville, MD 20850, USA.
          Tel: 301-838-3523
          Fax: 301-838-0208
          Email: cdtown@tigr.org
          DNA is from a doubled haploid provided by Tom Osborn.
          Seq primer: TR
          Class: sheared ends.
FEATURES             Location/Qualifiers
     source           1..609
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGLI82"
                     /clone_lib="BOGL"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT  157 a 120 c 140 g 192 t
ORIGIN
Query Match      34.6%; Score 18; DB 17; Length 609;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      28  CACTTGCGTAGGTATCT 45
          |||
Db      555 CACTTGCGTAGGTATCT 572

RESULT 12
BF519194/c
LOCUS   BF519194
DEFINITION EST456655 DSIL Medicago truncatula cDNA clone pBSIL-19P16, mRNA
sequence.
ACCESSION BF519194
VERSION   BF519194.1 GI:11607877
KEYWORDS  EST.
SOURCE    barrel medic.
ORGANISM  Medicago truncatula
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
          Medicago.
REFERENCE 1 (bases 1 to 614)
          Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S., Peng
          ,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S.,
          Holt,I.E. and Fraser,C.M.
          ESTs from leaves of Medicago truncatula after inoculation with

QY      13  BQ141233/c
          |||
Db      591 TGTGAGTGCACCTTGCAGT 574

RESULT 13
BQ141233/c
LOCUS   BQ141233
DEFINITION NF017D09PHI1078 Phoma-infected Medicago truncatula cDNA clone
sequence.
ACCESSION BQ141233
VERSION   BQ141233.1 GI:20277359
KEYWORDS  EST.
SOURCE    barrel medic.
ORGANISM  Medicago truncatula
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
          Medicago.
REFERENCE 1 (bases 1 to 634)
          Watson,B.S., Shin,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R.,
          Gonzales,R.A., Bell,C.J., Imman,J.T., Waugh,M.E., Sullivan,J.P.,
          May,G.D. and Paiva,N.L.
          Expressed Sequence Tags from the Samuel Roberts Noble Foundation
          Medicago truncatula phoma-infected library
          unpublished (2002)
          Contact: Paiva NL
          Plant Biology Division
          The Samuel Roberts Noble Foundation
          2510 Sam Noble Parkway, Ardmore, OK 73402, USA
          Tel: 580 221 7317

```

Colletotrichum trifolii
Unpublished (2000)
Contact: Deborah A. Samac
Department of Plant Pathology
University of Minnesota
495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
Tel: 612 625 1243
Fax: 651 649 5058
Email: debbys@puccini.crl.umn.edu
University of Minnesota name: M276624e
TIGR sequence name: MTFCD92TK
More information is available at: <http://chrystie.tamu.edu/medicago>
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

FEATURES Location/Qualifiers
 source 1..614
 /organism="Medicago truncatula"
 /cultivar="genotype Al7"
 /db_xref="taxon:3880"
 /clone="pDSIL-19P16"
 /clone_lib="DSIL"
 /tissue.type="leaves infected with Colletotrichum
 trifolii"
 /dev_stage="cotyledons and primary leaves harvested 5 and
 8 days after inoculation with Colletotrichum trifolii"
 /lab_host="E. coli strain XLOLR"
 /note="Vector: pBluescript SK-; Site_1: EORI; Site_2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from
 cotyledons and primary leaves harvested 5 and 8 days after
 inoculation with Colletotrichum trifolii. The cDNA was
 directionally ligated into the Uni-ZAP XR vector from
 Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-ZAP phage using Ex-Assist
 helper phage and propagated in XLOLR cells. Note: EST may
 be of fungal origin."
BASE COUNT 177 a 137 c 112 g 188 t
ORIGIN

Query Match 34.6%; Score 18; DB 12; Length 614;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 20 TGTGAGTGCACCTTGCAGT 37
 |||
Db 591 TGTGAGTGCACCTTGCAGT 574

RESULT 13
BQ141233/c
LOCUS BQ141233
DEFINITION NF017D09PHI1078 Phoma-infected Medicago truncatula cDNA clone
sequence.

ACCESSION BQ141233
VERSION BQ141233.1 GI:20277359
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE 1 (bases 1 to 634)
AUTHORS Watson,B.S., Shin,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R.,
Gonzales,R.A., Bell,C.J., Imman,J.T., Waugh,M.E., Sullivan,J.P.,
May,G.D. and Paiva,N.L.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation
JOURNAL Medicago truncatula phoma-infected library
COMMENT Unpublished (2002)
Contact: Paiva NL

Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7317

Fax: 580 221 7380
Email: nlpaiva@noble.org
Insert Length: 634 Std Error: 0.00
Plate: 017 row: D column: 09
Seq primer: TCACACAGGAAACAGCTATGAC.

FEATURES

Location/Qualifiers
1. .634
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF017D09PH"
/clone_lib="Phoma-infected"
/tissue_type="leaf"
/dev_stage="Pathogen-induced, young trifoliolate"

/note="Vector: pBluescript SK(-); Young trifoliolate leaves of Medicago truncatula were excised and dip-inoculated in a spore suspension of Phoma medicaginis, and incubated in humid dishes. Pools of leaves were harvested at 0, 15, and 30 minutes and 1, 2, 3, 6, 14, 24, 48, 72, and 96 hours, and used to prepare total RNA. cDNA was prepared from polyA+ enriched, pooled samples of equivalent amounts of total RNA from each sample. The cDNA was directionally ligated into the Uni-Zap XR vector (Stratagene) and packaged using the Gigapack III Gold packaging extracts. Phagemids containing cDNA inserts were in vivo excised from the recombinant Uni-Zap XR vector using ExAssist helper phage and the E. coli strain XL1-Blue MRF' (Stratagene). Excised plasmids were plated using SOLR cells."

BASE COUNT 184 a 140 c 119 g 189 t 2 others

ORIGIN

Query Match 34.6%; Score 18; DB 14; Length 634;

Best Local Similarity 100.0%; Pred. No. 31;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TGTGAGGTCACCTGCAGT 37

|||||

Db 599 TGTGAGGTCACCTGCAGT 582

RESULT 14

AW348152

LOCUS AW348152 645 bp mRNA linear EST 04-OCT-2000
DEFINITION GM210001A21C5 Gm-r1021 Glycine max cDNA clone Gm-r1021-106 3', mRNA sequence.

ACCESSION AW348152

VERSION AW348152.1 GI:6845862

KEYWORDS EST.

SOURCE soybean.

ORGANISM

Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

1 (bases 1 to 645)

Vodkin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Coryell, V., Erpelting, J., Raph, C., Shoop, E., Pardinias, J., Liu, L. and Lewin, H. A Functional Genomics Program for Soybean (NSF 9872565)
Unpublished (1999)

Other_ESTs: A1443992

Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
University of Illinois

Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582

Email: l-vodkin@uiuc.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genome

ystems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES

Source

Location/Qualifiers
1. .645
/organism="Glycine max"
/cultivar="Williams"
/db_xref="taxon:3847"
/clone="Gm-r1021-106"
/clone_lib="Gm-r1021"
/tissue_type="root"
/lab_host="XL10-Gold"

/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; Library Gm-r1021 is a sequence-driven, reracked set of the original library Gm-cl004 which was prepared from root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. Stratagene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. The Gm-cl004 library was constructed by Dr. Paul Keim & Virginia H. Coryell, Department of Biology, Box 5640, Northern Arizona University, Flagstaff, AZ 86011, email: paul.keim@uau.edu, virginia.coryell@uau.edu. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Computational Biology Centers, University of Minnesota,
http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
. Reracking was performed by Genome Systems, St. Louis,
http://www.genomesystems.com, and sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois,
http://www.life.uiuc.edu/biotech/keck.html."

BASE COUNT 198 a 129 c 132 g 170 t 16 others

ORIGIN

Query Match 34.6%; Score 18; DB 10; Length 645;

Best Local Similarity 100.0%; Pred. No. 31;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 TTGTGAGGTCACCTGCAG 36

|||||

Db 363 TTGTGAGGTCACCTGCAG 380

RESULT 15

BH680500/c

LOCUS BH680500

DEFINITION BOMG22TF BO_2_3_KB Brassica oleracea genomic clone BOMG22, DNA sequence.

ACCESSION BH680500

VERSION BH680500.1 GI:18750943

KEYWORDS GSS.

SOURCE Brassica oleracea.

ORGANISM

Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 875)

Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M. Whole genome shotgun sequencing of Brassica oleracea Unpublished (2001)

COMMENT Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

1. .875

/organism="Brassica oleracea"

/strain="TO1000DH3"

```
/db_xref="taxon:3712"
/clone="BOMGG22"
/clone_lib="BO_2_3_KB"
/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      281 a      211 c      165 g      218 t
ORIGIN
Query Match      34.6%; Score 18; DB 17; Length 875;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 28 CACTTGCAGTAGGTATCT 45
    |||||||
Db 52 CACTTGCAGTAGGTATCT 35
Search completed: November 26, 2002, 16:06:31
Job time : 2203 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 11:13:09 ; Search time 297 seconds
(without alignments)
394.289 Million cell updates/sec

Title: US-09-875-945-3
Perfect score: 52
Sequence: 1 ctcttcgcattgcatattt.....gcagtagtatctgtgcaca 52

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	100.0	52	AA140729	Human IRS-2 gene r
2	16	30.8	1134	AAK72383	Human immune/haema
3	16	30.8	1134	AAK72384	Human immune/haema
C 4	16	30.8	3074	AAZ96301	S. pneumoniae deri
C 5	16	30.8	3074	AAV42985	Streptococcus pneu
C 6	16	30.8	7995	AAV44507	Human LEXTI DNA cl
7	16	30.8	9244	AAV52201	Streptococcus pneu
8	16	30.8	13518	ABL03288	Drosophila melanog
C 9	15	28.8	261	AAV89473	EST clone CO223.

10	15	28.8	477	22	AA181087	Human polynucleoti
11	15	28.8	549	21	AAC93631	Cat flea head and
C 12	15	28.8	581	21	AAC09421	Human secreted pro
C 13	15	28.8	736	19	AAV44299	Human secreted pro
C 14	15	28.8	736	22	AAF98473	Human CDNA clone C
15	15	28.8	1038	23	ABL04509	Drosophila melanog
16	15	28.8	1117	23	ABL04505	Drosophila melanog
17	15	28.8	1172	23	ABL04525	Drosophila melanog
C 18	15	28.8	1356	22	ABA89156	Escherichia coli p
C 19	15	28.8	1381	23	ABL07871	Drosophila melanog
20	15	28.8	1422	22	AAH67155	C glutamicum codin
21	15	28.8	1545	22	AAF71625	Corynebacterium g1
C 22	15	28.8	2890	23	ABL10152	Drosophila melanog
C 23	15	28.8	3170	23	ABL04504	Drosophila melanog
C 24	15	28.8	3170	23	ABL04508	Drosophila melanog
C 25	15	28.8	3172	23	ABL04524	Drosophila melanog
C 26	15	28.8	3535	23	AA579443	DNA encoding novel
C 27	15	28.8	3958	23	ABL07870	Drosophila melanog
C 28	15	28.8	32222	22	AA534734	Human DNA for a no
29	15	28.8	37443	22	AA165592	Nucleotide sequenc
30	15	28.8	48254	22	ABA89141	Escherichia coli p
31	15	28.8	48345	22	ABA89142	Escherichia coli p
C 32	15	28.8	74962	22	AAAD15256	Human phosphatase
C 33	15	28.8	105184	24	ABK24122	Bacterial artific
C 34	15	28.8	198285	24	ABK84699	Human CDNA differe
C 35	15	28.8	198285	24	ABN97319	Gene #3817 used to
C 36	15	28.8	349980	22	AAH68530	C glutamicum codin
C 37	15	28.8	349980	22	AAH68531	C glutamicum codin
C 38	14	26.9	106	22	ABAL1470	Human nervous syst
C 39	14	26.9	249	22	ABAL1592	Human nervous syst
C 40	14	26.9	298	24	ABL75468	Corn tassell-derive
C 41	14	26.9	344	19	AAV52485	Streptococcus pneu
C 42	14	26.9	354	22	AAF64723	Novel human polynu
C 43	14	26.9	380	21	AAC29168	Human secreted pro
44	14	26.9	427	22	ABA09350	Human secreted pro
C 45	14	26.9	441	24	ABN96437	Gene #2935 used to

ALIGNMENTS

RESULT 1
AAL40729
ID AAL40729 standard; DNA; 52 BP.
XX
AC AAL40729;
XX
DT 25-SEP-2002 (first entry)
XX
DE Human IRS-2 gene related oligonucleotide 5956, SEQ ID No 3.
DE
DE Antidiabetic; anorectic; human insulin regulated substance; IRS-2 gene;
KW diabetes; obesity; metabolic disorder; ds.
XX
OS Homo sapiens.
XX
PN WO200194410-A1.
XX
PD 13-DEC-2001.
XX
PF 08-JUN-2001; 2001WO-SE01308.
XX
PR 08-JUN-2000; 2000US-210207P.
PR 09-JUN-2000; 2000SE-0002189.
XX
XX (METC-) METCON MEDICIN AB.
XX
XX Smith U; .
XX
XX WPI; 2002-566443/60.
XX
XX New non-coding sequences isolated upstream of human insulin regulated
PT substance-2 gene, useful as marker for predicting, diagnosing and

PT treating metabolic diseases or disorders e.g. obesity and diabetes -
XX Claim 1: Fig 5; 38pp; English.
PS
XX
CC The invention relates to an isolated, substantially purified nucleotide
CC sequence which is a non-coding sequence upstream of human insulin
CC regulated substance (IRS-2) gene. The isolated IRS-2 gene nucleic acid
CC region or its homologue is useful as a marker for insulin regulating
CC action in an assay for evaluating or screening substances for insulin
CC regulating properties in vitro, where adipocytes, hepatic cells, muscle
CC tissue cells or pancreatic cells are used as model cells. The isolated
CC IRS-2 gene nucleic acid region, its homologue, an IRS-2 transcript,
CC sequence information derived from an IRS-2 transcript, or a substance
CC having insulin regulated properties is useful for manufacture of a
CC medicament for treating diabetes and obesity, and as a veterinary
CC preparation. The isolated IRS-2 gene nucleic acid region is also useful
CC for diagnosis of IRS-2 related metabolic disorders, diabetes and/or
CC differentiating between various types or stages of the disorders, and for
CC determining if a patient in need of treatment with an insulin regulating
CC substance has the predisposition to respond to the treatment, comprising
CC measuring activation of IRS-2 by determining the amount or relative
CC increase/decrease of the IRS-2 protein or corresponding mRNA, where the
CC isolated IRS-2 gene nucleic acid region is used as a marker when
CC administering IRS to a sample of cells such as blood, adipocyte, muscle
CC or liver cells taken from the patient. This polynucleotide sequence
CC represents an isolated nucleic acid which is a non-coding sequence
CC upstream of the human insulin regulated substance (IRS-2) gene of the
CC invention.
XX
SQ Sequence 52 BP; 10 A; 13 C; 12 G; 17 T; 0 other;

Query Match 100.0%; Score 52; DB 24; Length 52;
Best Local Similarity 100.0%; Pred. No. 3.9e-20;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCCTCGCATGCCATATTTGTGAGGTCACTGTCAGTAGGATCTGTGCACA 52
|||||
Db 1 CTCCTCGCATGCCATATTTGTGAGGTCACTGTCAGTAGGATCTGTGCACA 52

RESULT 2
AAK72383
ID AAK72383 standard; DNA; 1134 BP.
XX
AC AAK72383;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27195.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PR (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX Disclosure; SEQ ID NO 27195; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)

CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 1134 BP; 319 A; 203 C; 188 G; 424 T; 0 other;

Query Match 30.8%; Score 16; DB 22; Length 1134;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 TGAGGTCACCTTGCAGT 37
Db 114 TGAGGTCACCTTGCAGT 129
|||||

RESULT 3
AAK72384
ID AAK72384 standard; DNA; 1134 BP.
XX
AC AAK72384;
XX
XX 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27196.
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytotstatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
XX WO200157182-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 17-JAN-2001; 2001WO-US01354;
XX
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 07-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 11-JUL-2000; 2000US-0217496.
XX PR 14-JUL-2000; 2000US-0218290.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 26-JUL-2000; 2000US-0220964.
XX PR 14-AUG-2000; 2000US-0224518.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0225213.
XX PR 14-AUG-2000; 2000US-0225214.
XX PR 14-AUG-2000; 2000US-0225256.
XX PR 14-AUG-2000; 2000US-0225267.
XX PR 14-AUG-2000; 2000US-0225268.
XX PR 14-AUG-2000; 2000US-0225270.
XX PR 14-AUG-2000; 2000US-0225270.
XX PR 14-AUG-2000; 2000US-0225447.
XX PR 14-AUG-2000; 2000US-0225757.
XX PR 14-AUG-2000; 2000US-0225758.
XX PR 14-AUG-2000; 2000US-0225759.
XX PR 18-AUG-2000; 2000US-0226279.
XX PR 22-AUG-2000; 2000US-0226681.

PR	22-AUG-2000;	2000US-0226868.	PR	17-NOV-2000;	2000US-0249208.
PR	22-AUG-2000;	2000US-0227182.	PR	17-NOV-2000;	2000US-0249209.
PR	23-AUG-2000;	2000US-0227009.	PR	17-NOV-2000;	2000US-0249210.
PR	30-AUG-2000;	2000US-0228924.	PR	17-NOV-2000;	2000US-0249211.
PR	01-SEP-2000;	2000US-0229287.	PR	17-NOV-2000;	2000US-0249212.
PR	01-SEP-2000;	2000US-0229343.	PR	17-NOV-2000;	2000US-0249213.
PR	01-SEP-2000;	2000US-0229344.	PR	17-NOV-2000;	2000US-0249214.
PR	01-SEP-2000;	2000US-0229345.	PR	17-NOV-2000;	2000US-0249215.
PR	03-SEP-2000;	2000US-0229509.	PR	17-NOV-2000;	2000US-0249216.
PR	03-SEP-2000;	2000US-0229513.	PR	17-NOV-2000;	2000US-0249217.
PR	06-SEP-2000;	2000US-0230437.	PR	17-NOV-2000;	2000US-0249218.
PR	06-SEP-2000;	2000US-0230438.	PR	17-NOV-2000;	2000US-0249244.
PR	08-SEP-2000;	2000US-0231242.	PR	17-NOV-2000;	2000US-0249245.
PR	08-SEP-2000;	2000US-0231243.	PR	17-NOV-2000;	2000US-0249264.
PR	08-SEP-2000;	2000US-0231244.	PR	17-NOV-2000;	2000US-0249265.
PR	08-SEP-2000;	2000US-0231143.	PR	17-NOV-2000;	2000US-0249297.
PR	08-SEP-2000;	2000US-0231414.	PR	17-NOV-2000;	2000US-0249299.
PR	08-SEP-2000;	2000US-0232080.	PR	17-NOV-2000;	2000US-0249300.
PR	08-SEP-2000;	2000US-0232081.	PR	01-DEC-2000;	2000US-0250160.
PR	12-SEP-2000;	2000US-0231968.	PR	01-DEC-2000;	2000US-0250391.
PR	14-SEP-2000;	2000US-0232397.	PR	05-DEC-2000;	2000US-0251030.
PR	14-SEP-2000;	2000US-0232398.	PR	05-DEC-2000;	2000US-0251988.
PR	14-SEP-2000;	2000US-0232399.	PR	05-DEC-2000;	2000US-0256719.
PR	14-SEP-2000;	2000US-0232400.	PR	06-DEC-2000;	2000US-0251479.
PR	14-SEP-2000;	2000US-0232401.	PR	08-DEC-2000;	2000US-0251856.
PR	14-SEP-2000;	2000US-0233063.	PR	08-DEC-2000;	2000US-0251868.
PR	14-SEP-2000;	2000US-0233064.	PR	08-DEC-2000;	2000US-0251869.
PR	14-SEP-2000;	2000US-0233065.	PR	08-DEC-2000;	2000US-0251989.
PR	21-SEP-2000;	2000US-0234223.	PR	08-DEC-2000;	2000US-0251990.
PR	21-SEP-2000;	2000US-0234274.	PR	11-DEC-2000;	2000US-0254097.
PR	25-SEP-2000;	2000US-0234997.	PR	03-JAN-2001;	2001US-0259678.
PR	25-SEP-2000;	2000US-0234998.	XX		
PR	26-SEP-2000;	2000US-0235484.	PA	(HUMA-) HUMAN GENOME SCI INC.	
PR	27-SEP-2000;	2000US-0235834.	XX		
PR	27-SEP-2000;	2000US-0235836.	XX	Rosen CA, Barash SC, Ruben SM;	
PR	29-SEP-2000;	2000US-0236327.	PI		
PR	29-SEP-2000;	2000US-0236367.	XX	WPI; 2001-483426/52.	
PR	29-SEP-2000;	2000US-0236368.	DR		
PR	29-SEP-2000;	2000US-0236369.	XX		
PR	29-SEP-2000;	2000US-0236370.	PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
PR	29-SEP-2000;	2000US-0236370.	PT	useful for preventing, diagnosing and/or treating cancers and	
PR	02-OCT-2000;	2000US-0236802.	PT	metastasis -	
PR	02-OCT-2000;	2000US-0237037.	XX		
PR	02-OCT-2000;	2000US-0237038.	XX	Disclosure; SEQ ID NO 27196; 3071pp + Sequence Listing; English.	
PR	02-OCT-2000;	2000US-0237039.	PS		
PR	02-OCT-2000;	2000US-0237039.	XX	AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)	
PR	13-OCT-2000;	2000US-0237040.	CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic	
PR	13-OCT-2000;	2000US-0239935.	CC	activity, and can be used in gene therapy and vaccine production. (I)	
PR	13-OCT-2000;	2000US-0239937.	CC	proteins and polynucleotides may be used in the prevention, diagnosis and	
PR	20-OCT-2000;	2000US-0240960.	CC	treatment of diseases associated with inappropriate (I) expression. For	
PR	20-OCT-2000;	2000US-0241785.	CC	example, they may be used to treat disorders associated with decreased	
PR	20-OCT-2000;	2000US-0241786.	CC	expression by rectifying mutations or deletions in a patient's genome	
PR	20-OCT-2000;	2000US-0241787.	CC	that affect the activity of (I) by expressing inactive proteins or to	
PR	20-OCT-2000;	2000US-0241808.	CC	supplement the patients own production of (I). Additionally, (I)	
PR	20-OCT-2000;	2000US-0241809.	CC	polynucleotides may be used to produce the secreted (I), by inserting	
PR	20-OCT-2000;	2000US-0241826.	CC	the nucleic acids into a host cell and culturing the cell to express the	
PR	01-NOV-2000;	2000US-0244617.	CC	protein. (I) proteins and polynucleotides may be used to prevent,	
PR	08-NOV-2000;	2000US-0246474.	CC	diagnose and treat immune/haematopoietic-related diseases, especially	
PR	08-NOV-2000;	2000US-0246475.	CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703	
PR	08-NOV-2000;	2000US-0246476.	CC	to AAK87694 represent human immune/haematopoietic antigen genomic	
PR	08-NOV-2000;	2000US-0246477.	CC	sequences from the present invention. AAK54942 to AAK54950 and AAM82169	
PR	08-NOV-2000;	2000US-0246478.	CC	represent sequences used in the exemplification of the present invention.	
PR	08-NOV-2000;	2000US-0246523.	XX		
PR	08-NOV-2000;	2000US-0246525.	XX	Sequence 1134 BP; 319 A; 203 C; 188 G; 424 T; 0 other;	
PR	08-NOV-2000;	2000US-0246526.			
PR	08-NOV-2000;	2000US-0246527.			
PR	08-NOV-2000;	2000US-0246528.			
PR	08-NOV-2000;	2000US-0246532.			
PR	08-NOV-2000;	2000US-0246609.			
PR	08-NOV-2000;	2000US-0246610.			
PR	08-NOV-2000;	2000US-0246611.			
PR	17-NOV-2000;	2000US-0249207.			
			Query Match 30.8%; Score 16; DB 22; Length 1134;		
			Best Local Similarity 100.0%; Pred. No. 15;		
			Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	22	TCAGGTCACCTTGCACT 37			
Db	114	TCAGGTCACCTTGCACT 129			

```
RESULT 4
AAZ96301/c
ID AAZ96301 standard; DNA; 3074 BP.
XX
AC AAZ96301;
XX
DT 10-APR-2000 (first entry)
XX
DE S. pneumoniae derived DNA from ORF #129.
XX
KW Treatment; prevention; disease; diagnosis; gene therapy; screening;
KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.
XX
OS Streptococcus pneumoniae.
XX
PN WO9806734-A1.
XX
PD 19-FEB-1998.
XX
PF 15-AUG-1997; 97WO-US14436.
XX
PR 16-AUG-1996; 96US-0024022.
XX
PA (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI Stodola RK;
XX
DR WPI: 1998-159452/14.
DR P-PSDB; AAY85942, AAY85943.
XX
PT Streptococcus pneumoniae proteins and related DNA - useful for
PT screening compounds for antibacterial activity
XX
PS Claim 4; Page 160-161; 640pp; English.
XX
CC This invention describes novel isolated Streptococcus pneumoniae
CC polynucleotides (see AAZ96173-Z96494) and their encoded proteins (see
CC AAY85792-Y86182). The DNA, vectors and host cells described in the
CC method of the invention are useful for the recombinant expression of the
CC polypeptides. The polypeptides are useful for treatment or prevention of
CC disease, or diagnosis of disease related to expression or activity of
CC such a polypeptide. They can also be used to screen for compounds which
CC interact with and inhibit or activate such a polypeptide. The
CC polypeptides (or DNA encoding them, via gene therapy) are also useful
CC for inducing an immunological response in a mammal. The antagonists are
CC useful to inhibit such bacterial polypeptides. The polypeptides are
CC particularly useful to identify antimicrobial compounds and antibiotics.
CC They are also useful to determine their role in pathogenesis of
CC infection, dysfunction and disease.
XX
SQ Sequence 3074 BP; 840 A; 666 C; 610 G; 958 T; 0 other;

Query Match 30.8%; Score 16; DB 19; Length 3074;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 CTGTCAGTAGGTATCT 45
Db 2562 CTGTCAGTAGGTATCT 2547

RESULT 5
AAV42985/c
ID AAV42985 standard; DNA; 3074 BP.
XX
AC AAV42985;
XX
DT 09-NOV-1998 (first entry)
XX
DE Streptococcus pneumoniae polypeptide coding region.
XX
KW Polypeptide; ORF; open reading frame; infection; bacterial;
```

```
KW streptococcal; bacteremia; diagnosis; prophylaxis; ds.
XX Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS complement (944..1777)
FT /*tag= a
FT /*note= "polypeptide"
FT complement (2323..2694)
FT /*tag= b
FT /*note= "polypeptide"
XX
PN WO9823631-A1.
XX
PD 04-JUN-1998.
XX
PF 24-NOV-1997; 97WO-US21976.
XX
PR 27-NOV-1996; 96US-0031879.
XX
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PA (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI Reid RH, Zarfos PN;
XX
DR WPI: 1998-322654/28.
DR P-PSDB; AAW62707, AAW62708.
XX
PT Streptococcus pneumoniae polynucleotides - useful for developing
PT products for diagnosis, prevention and treatment of infections e.g.
PT pneumonia, bacteremia, meningitis or endocarditis
XX
PS Claim 1; Page 89-90; 181pp; English.
XX
CC The sequence is that of a Streptococcal polypeptide coding region.
CC The polypeptide can potentially be used for the diagnosis and
CC prevention of bacterial infections, especially SP infection.
CC It may be used for the treatment of diseases such as otitis media,
CC conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural
CC empyema, endocarditis or infection of the cerebrospinal fluid.
XX
SQ Sequence 3074 BP; 841 A; 666 C; 609 G; 958 T; 0 other;

Query Match 30.8%; Score 16; DB 19; Length 3074;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 CTGTCAGTAGGTATCT 45
Db 2562 CTGTCAGTAGGTATCT 2547

RESULT 6
AAZ44507/c
ID AAZ44507 standard; DNA; 7995 BP.
XX
AC AAZ44507;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human LEKTI DNA clone CIT978SKB_94F21 contig 8, SPINK5 exon 6.
XX
KW Human; SPINK5; lympho-epithelial Kazal-type related inhibitor; LEKTI; ds;
KW serine protease inhibitor; atopic disease; Netherton's syndrome; asthma;
KW eczema; hayfever; antiallergic; antinflammatory;
KW dermatological; PCR primer; sequencing primer; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200164747-A1.
XX
PD 07-SEP-2001.
```

```

XX 02-MAR-2001; 2001WO-GB00897.
XX
XX 02-MAR-2000; 2000GB-0005098.
PR
XX 03-MAR-2000; 2000GB-0005229.
XX
XX (ISIS-) ISIS INNOVATION LTD.
PA
XX
XX Hovnanian A, Chavanas S, Cookson W, Moffat M, Walley A;
PI
XX
XX WPI; 2001-582149/65.
DR
XX
XX Determining susceptibility to atopic disease or carrier status of
PT Netherton's syndrome in humans by identifying variants of or mutations
PT in SPINK5, a gene encoding lympho-epithelial Kazal-type related
PT inhibitor
XX
XX Disclosure; Page 103-105; 123pp; English.
PS
XX Sequences AA544359-AA544514 represent the SPINK5 gene, contigs and
CC fragments of a SPINK5 clone, sequencing primers and PCR primers for
CC SPINK5. SPINK5 encodes lympho-epithelial kazal-type related inhibitor
CC (LEKTI), a serine protease inhibitor. Susceptibility or predisposition to
CC an atopic disease in a human subject can be detected by screening the
CC genome for one or more polymorphic variants of SPINK5 gene and/or
CC expression of a variant LEKTI protein in a tissue. Carrier status of a
CC subject or development of Netherton's syndrome is diagnosed by screening
CC for the presence of loss-of-function mutations in the SPINK5 gene. An
CC expression vector comprising a nucleic acid encoding a serine protease
CC inhibitor or its functional fragment can be used in screening for
CC compounds with potential pharmacological activity by determining the
CC serine protease activity of a protein previously identified as a ligand
CC of the LEKTI protein. The atopic diseases include Netherton's Syndrome,
CC asthma, eczema and hayfever.
XX
XX Sequence 7995 BP; 2543 A; 1244 C; 1217 G; 2989 T; 2 other;
SQ
Query Match 30.8%; Score 16; DB 22; Length 7995;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 15 ATATTTCGAGTCAC 30
Db 3347 ATATTTCGAGTCAC 3332
RESULT 7
AAV52201
ID AAV52201 standard; DNA; 9244 BP.
XX
XX AAV52201;
AC
XX
XX 23-OCT-1998 (first entry)
DT
XX
XX Streptococcus pneumoniae genome fragment SEQ ID NO:68.
DE
XX
XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
XX
XX Streptococcus pneumoniae.
OS
XX WO9818931-A2.
PN
XX
XX 07-MAY-1998.
PD
XX
XX 30-OCT-1997; 97WO-US15988.
PF
XX
XX 31-OCT-1996; 96US-0029960.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI Kunsch CA, Rosen CA;
PI

```

```

XX WPI; 1998-272225/24.
XX
XX Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
XX
XX Claim 1; Page 556-561; 1409pp; English.
PS
XX
XX The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC recorded on it, or a representative fragment or a sequence at least 95%
CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC Streptococcus pneumoniae. The present invention also describes an
CC isolated nucleic acid molecule encoding a homologue of any of the
CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridise to the target sequence and
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the S. pneumoniae genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the S. pneumoniae genome of commercial importance, or
CC expression modulating fragments of the S. pneumoniae genome. Products
CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for S. pneumoniae.
XX
XX Sequence 9244 BP; 2718 A; 1876 C; 1862 G; 2788 T; 0 other;
SQ
Query Match 30.8%; Score 16; DB 19; Length 9244;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 30 CTTGCAGTAGGTATCT 45
Db 755 CTTGCAGTAGGTATCT 770
RESULT 8
ABL03288
ID ABL03288 standard; cDNA; 13518 BP.
XX
XX ABL03288;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 4346.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
PI

```

DR P-PSDB; ABB59185.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 4346; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
 CC sequences (ABLI01840-ABLI16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 13518 BP; 3688 A; 3113 C; 3012 G; 3705 T; 0 other;
 Query Match 30.8%; Score 16; DB 23; Length 13518;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 ATTGCCATATTGTGA 24
 Db 6155 ATTGCCATATTGTGA 6170
 |||||||||||||||
 RESULT 9
 AAV89473/c
 ID AAV89473 standard; cDNA: 261 BP.
 XX
 AC AAV89473;
 XX
 DT 15-FEB-1999 (first entry)
 XX
 DE EST clone CO223.
 XX
 KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
 KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
 KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
 KW gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9845436-A2.
 XX
 PD 15-OCT-1998.
 XX
 PF 10-APR-1998; 98WO-US06955.
 XX
 PR 10-APR-1997; 97US-0838821.
 XX
 PA (GEMY) GENETICS INST INC.
 XX
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 PI Racie LA, Spaulding V, Treacy M,
 XX
 DR WPI; 1999-070077/06.
 XX
 PT New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries.
 XX
 PS Claim 1; Page 229; 618pp; English.
 XX
 CC The present sequence represents a human expressed sequence tag (EST).
 CC The polynucleotide, which is a secreted EST, and the encoded protein
 CC are predicted to have useful biological activities which would make
 CC them suitable for treating, preventing or ameliorating medical

CC conditions in humans and animals, although no supporting data is
 CC given. Suggested activities include nutritional activity, immune
 CC stimulating or suppressing activity, haematopoiesis regulating
 CC activity, tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
 CC activity, receptor/ligand activity, anti-inflammatory activity,
 CC cadherin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The polynucleotide may also be useful for gene therapy.
 XX
 SQ Sequence 261 BP; 47 A; 80 C; 49 G; 85 T; 0 other;
 Query Match 28.8%; Score 15; DB 20; Length 261;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 37 TAGGTATCTGTGCAC 51
 Db 208 TAGGTATCTGTGCAC 194
 |||||||||||||||
 RESULT 10
 AAI81087
 ID AAI81087 standard; cDNA: 477 BP.
 XX
 AC AAI81087;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 1147.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US04927.
 XX
 PR 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 PI
 XX
 DR WPI; 2001-514838/56.
 DR P-PSDB; AAO01156.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 XX
 PS Claim 1; SEQ ID NO 1147; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

```

XX
SQ      Sequence 549 BP; 184 A; 96 C; 103 G; 166 T; 0 other;

Query Match      28.8%; Score 15; DB 21; Length 549;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 ATTGCCATATTGTG 23
      |||||||
DB      63 ATTGCCATATTGTG 77

RESULT 12
AAC09421/C
ID      AAC09421 standard; cDNA; 581 BP.
XX      AC
XX      AAC09421;
DT      06-OCT-2000 (first entry)
XX      DE
XX      Human secreted protein 5' EST, SEQ ID NO: 13496.
XX      Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW      gene therapy; chromosome mapping; ss.
XX      OS Homo sapiens.
XX      EP1033401-A2.
XX      PN
XX      06-SEP-2000.
XX      PD
XX      21-FEB-2000; 2000EP-0200610.
XX      PF
XX      26-FEB-1999; 99US-0122487.
XX      PR
XX      (GEST ) GENSET.
XX      PA
XX      Dumas Milne Edwards J, Duclert A, Giordano J;
XX      WPI; 2000-500381/45.
XX      PT
XX      New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT      obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT      diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX      Claim 1; SEQ ID 13496; 71pp + CD-ROM; English.
XX      PS
XX      The present sequence is one of a large number of 5' ESTs derived from
CC      mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC      identified within the present sequence. The 5' ESTs were prepared from
CC      total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC      sequences usually correspond mainly to the 3' untranslated region (UTR)
CC      of the mRNA because they are often obtained from oligo-dT primed cDNA
CC      libraries. Such ESTs are not well suited for isolating cDNA sequences
CC      derived from the 5' ends of mRNAs and even in those cases where longer
CC      cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC      5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC      used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC      in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC      They are used to obtain upstream regulatory sequences and to design
CC      expression and secretion vectors.
XX      SQ
      Sequence 581' BP; 203 A; 115 C; 104 G; 157 T; 2 other;

Query Match      28.8%; Score 15; DB 21; Length 581;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      24 AGGTCACCTGTCAGTA 38
      |||||||
DB      86 AGGTCACCTGTCAGTA 72

```

```

RESULT 13
AAV44299/c
ID AAV44299 standard; cDNA; 736 BP.
XX
AC AAV44299;
XX
DT 06-OCT-1998 (first entry)
XX
DE Human secreted protein clone C0223_3 cDNA.
XX
KW Secreted protein; human adult brain; nutrition; cytokine; stimulant;
KW cell proliferation; differentiation; immune system; suppressor; ligand;
KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
KW chemotaxis; chemokinetic; thrombosis; receptor; cadherin; tumour;
KW anti-inflammatory; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 176..520
FT /tag= a
FT /product= secreted protein
FT /note= "isolated from clone C0223_3"
XX
PN WO9827205-A2.
XX
PD 25-JUN-1998.
XX
PF 17-DEC-1997; 97WO-US23330.
XX
PR 16-DEC-1997; 97US-0991872.
PR 18-DEC-1996; 96US-0769192.
PR 13-JAN-1997; 97US-0783401.
XX
XX (GEMV ) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1998-362774/31.
DR P-PSDB; AAW64224.
XX
XX New polynucleotides and secreted proteins - obtained from human
XX foecal brain, human adult testes, human adult brain and human adult
XX salivary gland cDNA libraries
XX
PS Claim 26a; Page 78-79; 110pp; English.
XX
CC This sequence encodes a novel secreted protein from clone C0223_3
CC isolated from a human adult brain cDNA library. This protein has
CC applications for nutritional use, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or
CC suppressing activity, hematopoiesis regulating activity, tissue growth
CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
CC haemostatic and thrombotic activity, receptor/ligand activity,
CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC tumour inhibition activity and other activities.
XX
SQ Sequence 736 BP; 236 A; 180 C; 96 G; 221 T; 3 other;

Query Match 28.8%; Score 15; DB 19; Length 736;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 TAGGTATCTGTGCAC 51
DB 326 TAGGTATCTGTGCAC 312

RESULT 14
AAV98473/c
ID AAV98473 standard; cDNA; 736 BP.
XX

```

```

AC AAF98473;
XX
DT 07-JUN-2001 (first entry)
XX
DE Human cDNA clone C0233_3 sequence SEQ ID 167.
XX
KW Human; secreted protein; nutrient; cytokine modulator; proliferation;
KW differentiation; immune system modulator; tissue growth; chemotactic;
KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition; ss;
KW haematopoiesis.
XX
OS Homo sapiens.
XX
PN WO200119988-A1.
XX
PD 22-MAR-2001.
XX
PF 14-SEP-2000; 2000WO-US25135.
XX
PR 17-SEP-1999; 99US-0398829.
XX
PA (GEMV ) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Bowman MR, Spaulding V, Agostino MJ;
XX
DR WPI; 2001-244801/25.
DR P-PSDB; AAB90734.
XX
XX Isolated nucleic acids encoding polypeptides, useful for modulating
XX e.g. cytokine and cell proliferation/differentiation activity, the
XX immune system and hematopoiesis regulating activity -
XX
PS Disclosure; Page 493; 557pp; English.
XX
CC Human cDNA clones represented in AAF98374 - AAF98489 encode secreted
CC proteins AAB90667 - AAB90750. The cDNA clones are isolated from various
CC tissue types, and may be used in the prevention, treatment and diagnosis
CC of diseases associated with inappropriate protein expression. The
CC polypeptides and nucleic acids may be used as nutrients or to modulate
CC cytokine and cell proliferation/differentiation activity and may also be
CC involved in modulation of the immune system. The cDNA sequences,
CC proteins, their agonists and/or antagonists exhibit haematopoiesis
CC regulating activity; tissue growth activity; activin/inhibin activity;
CC chemotactic/chemokinetic activity; haemostatic and thrombolytic
CC activity; receptor/ligand activity; anti-inflammatory activity;
CC haematopoiesis activity; cadherin/tumour suppressor activity; and/or
CC tumour inhibition activity. Included in the invention are probes
CC represented in AAF98490 - AAF98572 which are specific for the cDNA clones
CC encoding the secreted proteins.
XX
SQ Sequence 736 BP; 237 A; 180 C; 96 G; 221 T; 2 other;

Query Match 28.8%; Score 15; DB 22; Length 736;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 TAGGTATCTGTGCAC 51
DB 326 TAGGTATCTGTGCAC 312

RESULT 15
ABL04509
ID ABL04509 standard; cDNA; 1038 BP.
XX
AC ABL04509;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 8009.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;

```

KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
PF 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
XX
DR P-PSDB; ABB60406.
XX
XX
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX
PS Claim 1; SEQ ID NO 8009; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1038 BP; 315 A; 236 C; 240 G; 247 T; 0 other;

Query Match 28.8%; Score 15; DB 23; Length 1038;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCCTCGCATTGCCA 15
DB 519 CTCCTCGCATTGCCA 533
|||||

Search completed: November 26, 2002, 14:37:06
Job time : 302 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	22	42.3	207433	6	US-10-277-216-5	Sequence 5, Appli
2	21.8	41.9	1810	5	US-09-724-676-10671	Sequence 10671, A
3	21.8	41.9	1890	5	US-09-724-676-10672	Sequence 10672, A
C 4	21.6	41.5	2656	5	US-09-724-676-26198	Sequence 26198, A
C 5	21.6	41.5	2795	5	US-09-724-676-26197	Sequence 26197, A
6	21.4	41.2	202001	6	US-10-274-990-3	Sequence 3, Appli
C 7	20.4	39.2	4994	5	US-09-724-676-13078	Sequence 13078, A
C 8	20.4	39.2	5083	5	US-09-724-676-13072	Sequence 13072, A
C 9	20.4	39.2	5191	5	US-09-724-676-13077	Sequence 13077, A
C 10	20.4	39.2	5280	5	US-09-724-676-13071	Sequence 13071, A
C 11	20.4	39.2	5333	5	US-09-724-676-13076	Sequence 13076, A
C 12	20.4	39.2	5422	5	US-09-724-676-13070	Sequence 13070, A
C 13	20.4	39.2	5530	5	US-09-724-676-13075	Sequence 13075, A
C 14	20.4	39.2	5619	5	US-09-724-676-13069	Sequence 13069, A
C 15	20.4	39.2	5760	5	US-09-724-676-13093	Sequence 13093, A
C 16	20.4	39.2	5761	5	US-09-724-676-13043	Sequence 13040, A
C 17	20.4	39.2	5849	5	US-09-724-676-13087	Sequence 13087, A
C 18	20.4	39.2	5850	5	US-09-724-676-13034	Sequence 13034, A
C 19	20.4	39.2	5957	5	US-09-724-676-13092	Sequence 13092, A
C 20	20.4	39.2	5958	5	US-09-724-676-13039	Sequence 13039, A
C 21	20.4	39.2	6046	5	US-09-724-676-13086	Sequence 13086, A
C 22	20.4	39.2	6047	5	US-09-724-676-13033	Sequence 13033, A
C 23	20.4	39.2	6099	5	US-09-724-676-13091	Sequence 13091, A
C 24	20.4	39.2	6100	5	US-09-724-676-13038	Sequence 13038, A
C 25	20.4	39.2	6179	5	US-09-724-676-13063	Sequence 13063, A
C 26	20.4	39.2	6188	5	US-09-724-676-13085	Sequence 13085, A


```

RESULT 10
US-09-724-676-13071/c
; Sequence 13071, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: CompuGen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 CompuGen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2

```

RESULT 13
US-09-724-676-13075/c
; Sequence 13075, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing

Db 4485 CTCCTGTGCAAAATGGAAGAGGTCTGTATGGAGCAGATGCTGTGTC 4440

Search completed: November 26, 2002, 14:32:31
Job time : 116 secs

FILE REFERENCE: 129181.4 Compugen
CURRENT APPLICATION NUMBER: US/09/724,676
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 13075
LENGTH: 5530
TYPE: DNA
ORGANISM: Homo sapiens
US-09-724-676-13075

Query Match 39.2%; Score 20.4; DB 5; Length 5530;
Best Local Similarity 65.2%; Pred. No. 43;
Matches 30; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 4 CTCGCATTGCCATATTTGTGAGGTCACTTGCAGTAGGTATCTGTGC 49
Db 3719 CTCCTGTGCAAAATGGAAGAGGTCTGTATGGAGCAGATGCTGTGTC 3674

RESULT 14
US-09-724-676-13069/c
Sequence 13069, Application US/09724676
GENERAL INFORMATION:
APPLICANT: Compugen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Compugen
CURRENT APPLICATION NUMBER: US/09/724,676
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 13069
LENGTH: 5619
TYPE: DNA
ORGANISM: Homo sapiens
US-09-724-676-13069

Query Match 39.2%; Score 20.4; DB 5; Length 5619;
Best Local Similarity 65.2%; Pred. No. 43;
Matches 30; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 4 CTCGCATTGCCATATTTGTGAGGTCACTTGCAGTAGGTATCTGTGC 49
Db 3808 CTCCTGTGCAAAATGGAAGAGGTCTGTATGGAGCAGATGCTGTGTC 3763

RESULT 15
US-09-724-676-13093/c
Sequence 13093, Application US/09724676
GENERAL INFORMATION:
APPLICANT: Compugen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Compugen
CURRENT APPLICATION NUMBER: US/09/724,676
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 13093
LENGTH: 5760
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (14)..(14)
OTHER INFORMATION: n is a,c,g, or t
US-09-724-676-13093

Query Match 39.2%; Score 20.4; DB 5; Length 5760;
Best Local Similarity 65.2%; Pred. No. 43;
Matches 30; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 4 CTCGCATTGCCATATTTGTGAGGTCACTTGCAGTAGGTATCTGTGC 49

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 08:58:05 ; Search time 3781 Seconds
(without alignments)
345.784 Million cell updates/sec

Title: US-09-875-945-3
Perfect score: 52
Sequence: 1 ctctcgcattgcacatatt.....gcagtaggtatctgtgcaca 52

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues 49582208
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents_NA_Main:*

1: /cgn2_6/ptodata/1/pna/pctus_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US080_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US081_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US082_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US083_COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US084_COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US085_COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US086_COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US087_COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US088_COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US089_COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US090_COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US091_COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US092_COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US093_COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US094_COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US095A_COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US095B_COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US095C_COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US095D_COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US096A_COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US096B_COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US096C_COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US096D_COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US096E_COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US097A_COMB.seq.*
29: /cgn2_6/ptodata/1/pna/US097B_COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US097C_COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US098A_COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US098B_COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US098C_COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US099A_COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US099B_COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US099C_COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US099D_COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US100A_COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US100B_COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US101A_COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US101B_COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US102A_COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US102B_COMB.seq.*

44: /cgn2_6/ptodata/1/pna/US6000_COMB.seq.*
45: /cgn2_6/ptodata/1/pna/US6001_COMB.seq.*
46: /cgn2_6/ptodata/1/pna/US6002_COMB.seq.*
47: /cgn2_6/ptodata/1/pna/US6003_COMB.seq.*
48: /cgn2_6/ptodata/1/pna/US6004_COMB.seq.*
49: /cgn2_6/ptodata/1/pna/US6005_COMB.seq.*
50: /cgn2_6/ptodata/1/pna/US6006_COMB.seq.*
51: /cgn2_6/ptodata/1/pna/US6007_COMB.seq.*
52: /cgn2_6/ptodata/1/pna/US6008_COMB.seq.*
53: /cgn2_6/ptodata/1/pna/US6009_COMB.seq.*
54: /cgn2_6/ptodata/1/pna/US6010_COMB.seq.*
55: /cgn2_6/ptodata/1/pna/US6011_COMB.seq.*
56: /cgn2_6/ptodata/1/pna/US6012_COMB.seq.*
57: /cgn2_6/ptodata/1/pna/US6013_COMB.seq.*
58: /cgn2_6/ptodata/1/pna/US6014_COMB.seq.*
59: /cgn2_6/ptodata/1/pna/US6015_COMB.seq.*
60: /cgn2_6/ptodata/1/pna/US6016_COMB.seq.*
61: /cgn2_6/ptodata/1/pna/US6017_COMB.seq.*
62: /cgn2_6/ptodata/1/pna/US6018_COMB.seq.*
63: /cgn2_6/ptodata/1/pna/US6019_COMB.seq.*
64: /cgn2_6/ptodata/1/pna/US6020_COMB.seq.*
65: /cgn2_6/ptodata/1/pna/US6021_COMB.seq.*
66: /cgn2_6/ptodata/1/pna/US6022_COMB.seq.*
67: /cgn2_6/ptodata/1/pna/US6023_COMB.seq.*
68: /cgn2_6/ptodata/1/pna/US6024_COMB.seq.*
69: /cgn2_6/ptodata/1/pna/US6025_COMB.seq.*
70: /cgn2_6/ptodata/1/pna/US6026_COMB.seq.*
71: /cgn2_6/ptodata/1/pna/US6027_COMB.seq.*
72: /cgn2_6/ptodata/1/pna/US6028_COMB.seq.*
73: /cgn2_6/ptodata/1/pna/US6029_COMB.seq.*
74: /cgn2_6/ptodata/1/pna/US6030_COMB.seq.*
75: /cgn2_6/ptodata/1/pna/US6031_COMB.seq.*
76: /cgn2_6/ptodata/1/pna/US6032_COMB.seq.*
77: /cgn2_6/ptodata/1/pna/US6033_COMB.seq.*
78: /cgn2_6/ptodata/1/pna/US6034_COMB.seq.*
79: /cgn2_6/ptodata/1/pna/US6035_COMB.seq.*
80: /cgn2_6/ptodata/1/pna/US6036_COMB.seq.*
81: /cgn2_6/ptodata/1/pna/US6037_COMB.seq.*
82: /cgn2_6/ptodata/1/pna/US6038_COMB.seq.*
83: /cgn2_6/ptodata/1/pna/US6039_COMB.seq.*
84: /cgn2_6/ptodata/1/pna/US6040_COMB.seq.*
85: /cgn2_6/ptodata/1/pna/US6041_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	52	100.0	52	33	US-09-875-945-3
2	24	46.2	609	21	US-09-540-764-46736
3	23.8	45.8	454	20	US-09-534-844A-6724
4	23.8	45.8	471	19	US-09-505-532-21887
5	23.8	45.8	471	31	US-09-819-091A-21887
6	23.8	45.8	401	62	US-60-180-489-4236
7	23.8	45.8	1466	19	US-09-513-936A-13958
8	23.8	45.8	1466	24	US-09-620-394B-2192
9	23.8	45.8	97554	27	US-09-692-412-37
10	23.8	45.8	97554	31	US-09-803-736-1017
11	23.6	45.4	1251	31	US-09-815-264-14419
12	23.6	45.4	14677	24	US-09-620-392-38406
13	23.6	45.4	14677	28	US-09-702-134-27024
14	23.6	45.4	14677	31	US-09-815-264-74648
15	23.4	45.0	311	69	US-60-253-456-20031
16	23.4	45.0	312	69	US-60-253-456-19972
17	23.4	45.0	524	18	US-09-421-106-30259
18	23.4	45.0	524	19	US-09-521-640-298983
19	23.4	45.0	560	1	PCT-US01-00663-10435
20	23.4	45.0	560	33	US-09-864-761-7874
21	23.4	45.0	560	41	US-10-182-993-10130

22	23.4	45.0	560	41	US-10-102-995-7874	Sequence 7874, Ap
23	23.4	45.0	560	42	US-10-203-134-10591	Sequence 10591, A
24	23.4	45.0	560	42	US-10-203-135-10198	Sequence 10198, A
25	23.4	45.0	560	42	US-10-203-136-10725	Sequence 10725, A
26	23.4	45.0	560	42	US-10-203-137-10435	Sequence 10435, A
27	23.4	45.0	560	42	US-10-203-139-10130	Sequence 10130, A
28	23.4	45.0	194788	19	US-09-528-2378-1930	Sequence 1930, Ap
C 29	23.2	44.6	201	22	US-09-572-409-58720	Sequence 58720, A
	23.2	44.6	581	60	US-60-160-189-4846	Sequence 4846, Ap
C 30	23.2	44.6	582	60	US-60-160-189-3121	Sequence 3121, Ap
	23.2	44.6	582	60	US-60-253-456-18285	Sequence 18285, A
31	23.2	44.6	603	69	US-60-253-456-18285	Sequence 18285, A
32	23.2	44.6	603	69	US-09-505-532-26023	Sequence 26023, A
33	23.2	44.6	615	31	US-09-819-0918-26023	Sequence 26023, A
34	23.2	44.6	1647	38	US-09-708-427-28565	Sequence 28565, A
C 35	23.2	44.6	1647	25	US-09-938-8424-19	Sequence 19, Appl
	23.2	44.6	1647	35	US-09-614-150-37309	Sequence 37309, A
C 36	23.2	44.6	3899	23	US-60-191-637-36923	Sequence 36923, A
C 37	23.2	44.6	3899	63	US-60-229-518-182	Sequence 182, App
C 38	23.2	44.6	3899	63	US-09-534-859-627	Sequence 627, App
C 39	23.2	44.6	32768	66	US-09-803-736-627	Sequence 627, App
40	23.2	44.6	42446	31	US-09-528-2378-1304	Sequence 1304, Ap
41	23.2	44.6	42446	31	US-09-654-617-419666	Sequence 419666, A
42	23.2	44.6	179447	19	US-09-669-8176-34711	Sequence 34711, A
C 43	23	44.2	525	25	US-09-684-016-419666	Sequence 419666, A
C 44	23	44.2	525	26		
C 45	23	44.2	525	27		

US-09-692-412-37/C
; Sequence 37, Application US/09692412
; GENERAL INFORMATION:
; APPLICANT: Bush, David F.
; APPLICANT: Rounsley, Steven D.

[illegible]

Search completed: November 26, 2002, 12:46:22
Job time : 3826 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 08:58:04 ; Search time 2199 Seconds
(without alignments)
382.976 Million cell updates/sec

Title: US-09-875-945-3

Perfect score: 52

Sequence: 1 ctcctgcattgcatttt.....gcagtaggtatctgtgcaca 52

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*

1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_estc.*
9: gb_est1.*
10: gb_est2.*
11: gb_est3.*
12: gb_est4.*
13: gb_est5.*
14: em_estfun.*
15: em_estom.*
16: gb_gss.*
17: gb_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_other.*
25: em_gss_pro.*
26: em_gss_rod.*
27: em_gss_rod.*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	26.8	51.5	427	10	BB683792
c 2	26.8	51.5	970	17	B11776 F5M4-Sp6 IG
c 3	25.6	49.2	421	14	N71339 za3of03.s1
c 4	25.6	49.2	972	14	BQ964606 AGENCOURT
c 5	25.4	48.8	927	17	AL173557 Tetraodon
c 6	25.2	48.5	251	10	BB174493

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

c 7	25.2	48.5	300	10	BB146258
c 8	25.2	48.5	414	10	BB688721
c 9	25.2	48.5	445	14	BQ557088 H404A402-
c 10	25.2	48.5	522	13	B1133669 UI-M-BH3-
c 11	25.2	48.5	526	10	BB772002 BB772002
c 12	25.2	48.5	546	13	BM120065 L0934B10-
c 13	25.2	48.5	547	14	BM941987 UI-M-CG0P
c 14	25.2	48.5	554	10	AW540913 C0138F08-
c 15	25.2	48.5	590	13	BM239482 K0535B10-
c 16	25.2	48.5	610	10	BB770953 BB770953
c 17	25.2	48.5	636	13	BM240267 K0549E12-
c 18	25.2	48.5	725	13	BM119782 L0930E05-
c 19	25	48.1	518	10	BB759087 BB759087
c 20	24.8	47.7	379	10	AV805772 AV805772
c 21	24.6	47.3	502	10	BB757635 BB757635
c 22	24.6	47.3	925	17	B10024 F5M4-Sp6.1
c 23	24.4	46.9	523	17	AQ189253 HS_3205_B
c 24	24.4	46.9	588	17	AQ154359 HS_3006_A
c 25	24.4	46.9	1088	17	AL183251 Tetraodon
c 26	24.2	46.5	466	17	AZ405313 LM0174P02
c 27	24.2	46.5	554	9	AI791118 uk55h09.y
c 28	24	46.2	282	12	BF377027 OY0-PN008
c 29	24	46.2	660	14	BM957572 fy67g09.x
c 30	24	46.2	993	17	CNS00JRO
c 31	23.8	45.8	421	13	BM385808 UI-R-CN1
c 32	23.8	45.8	490	12	BE951240 UI-M-BH4-
c 33	23.8	45.8	612	10	AW121574 UI-M-BH2
c 34	23.8	45.8	661	10	AW035264 EST280526
c 35	23.6	45.4	257	10	BB178340 BB178340
c 36	23.6	45.4	278	10	BB239760 BB239760
c 37	23.6	45.4	290	10	BB376055 BB376055
c 38	23.6	45.4	315	10	AV309699 AV309699
c 39	23.6	45.4	334	10	BB224560 BB224560
c 40	23.6	45.4	463	10	BB783268 BB783268
c 41	23.6	45.4	471	12	BF092814 MR4-TN011
c 42	23.6	45.4	534	10	BB769187 BB769187
c 43	23.6	45.4	597	17	TSP458750
c 44	23.4	45.0	542	10	AV741437 AV741437
c 45	23.2	44.6	305	17	AQ100184 HS_3049_B

ALIGNMENTS

RESULT 1	BB683792/c	427 bp	mrna	linear	EST 10-OCT-2001
LOCUS	BB683792	RIKEN full-length enriched, 12 days embryo female			
DEFINITION	BB683792	mullerian duct Mus musculus cDNA clone 6820409B13 3', mRNA			
ACCESSION	BB683792	sequence.			
VERSION	BB683792.1	GI:16010525			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
	1 (bases 1 to 427)				
	Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hangaki,T.,				
	Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii				
	,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,				
	Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.				
	, Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,				
	Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa				
	,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.				
	, Wachihi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.				
	RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.				
	2001)				
TITLE	Unpublished (2001)				
JOURNAL	Contact: Yoshihide Hayashizaki				
COMMENT	Laboratory for Genome Exploration Research Group, RIKEN Genomic				
	Sciences Center(GSC), Yokohama Institute				
	The Institute of Physical and Chemical Research (RIKEN)				


```

FEATURES
source
Location/Qualifiers
1..421
/organism="Homo sapiens"
/db_xref="GDB:1238998"
/db_xref="taxon:9606"
/clone_lib="IMAGE:294077"
/clone_lib="Soares fetal liver spleen lNFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACGGGAAGAATAAATTAAGATCTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      130 a   73 c   100 g   113 t
ORIGIN
Query Match          49.2%; Score 25.6; DB 14; Length 421;
Best Local Similarity 69.4%; Pred. No. 35;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 3 CCTCGCATTCACATATTGTGAGGTGCACATTCGAGTAGGTATCTGTGCAC 51
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 CCTGCCATTCGATCCTGGGGAGAACTTCGNTAGTATGTGACCAC 366

RESULT 4
BQ964606
LOCUS
DEFINITION
AGENCOURT_10050390 NIH_MGC_134 Mus musculus cDNA clone
IMAGE:6513792 5', mRNA sequence.
ACCESSION
BQ964606
VERSION
BQ964606.1 GI:22380084
KEYWORDS
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 972)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14088 row: b column: 01
High quality sequence start: 19
High quality sequence stop: 660.
Features
Location/Qualifiers
1..972
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="IMAGE:6513792"
/clone_lib="NIH_MGC_134"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pCMV-SF0R6.1.cdb; Site_1: EcoRV; Site_2:
NotI; Cloned unidirectionally. Primer: Oligo dT. Average
insert size 1-7 kb. Constructed by ResGen, Invitrogen
Corp. Note: this is a NIH_MGC Library."
BASE COUNT      209 a   215 c   270 g   278 t
ORIGIN

FEATURES
source
Location/Qualifiers
1..927
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="219C10"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG219BB05LPl-end : T7"
BASE COUNT      266 a   196 c   196 g   266 t
ORIGIN
Query Match          48.8%; Score 25.4; DB 17; Length 927;
Best Local Similarity 74.4%; Pred. No. 55;
Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 8 CATTGCCATATTGTGAGGTTCACCTTCAGTAGGTATCTGTGCA 50
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 230 CACTGCCATTTTGGCGAGGTTCATTTCACCCAGATTCTGACCA 188

RESULT 6
BB174493/c
LOCUS
DEFINITION
musculus cDNA clone A230051P14 3', mRNA sequence.
ACCESSION
BB174493
VERSION
BB174493.1 GI:8833576

```


Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 445)
VanBuren,V., Piao,Y., Dudekula,D.B., Qian,Y., Carter,M.G., Martin
P.R., Stagg,C.A., Bassey,U., Alba,K., Hamatani,T., Kargul,G.J.,
Luo,A.G. and Ko,M.S.H.
Assembly, verification, and initial annotation of NIA 7.4K mouse
cDNA clone set
Unpublished (2002)
Other_Ests: H4044A02-5
Contact: Yong Qian
Laboratory of Genetics
National Institute on Aging/National Institutes of Health

COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
 Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.
 e mouse tissues.

FEATURES
source

Location/Qualifiers
 1..526
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="G370139F11"
 /clone_lib="RIKEN full-length enriched, B16 F10Y cells"
 /cell_type="B16 F10Y cells"
 /note="pooled tissues ; (tissue_type=cerebellum,
 dev_stage=16 days neonate, sex=mixed),
 (tissue_type=cerebellum, dev_stage=0 day neonate,
 sex=mixed), (tissue_type=hippocampus, dev_stage=adult,
 sex=male), (tissue_type=whole body, dev_stage=9 days
 embryo, sex=mixed), (tissue_type=lung, dev_stage=13 days
 embryo, sex=mixed)"
 145 a 121 c 137 g 123 t

BASE COUNT
ORIGIN

Query Match 48.5%; Score 25.2; DB 10; Length 526;
 Best Local Similarity 71.7%; Pred. No. 53;
 Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 7 GCATTGCCATATTTGTAGGTCACCTGCGAGTATCTGTGCACA 52
 || ||||| || ||||| ||||| ||||| ||||| ||||| ||
 Db 335 GCCTTGCCTTAATGTGACAGCAGCTCCAGCAGGTACCTGTGCCCA 290

RESULT 12
BM120065

LOCUS BM120065 546 bp mRNA linear EST 01-FEB-2002
 DEFINITION L0934B10-3 NTA Mouse Newborn Kidney cDNA Library (Long) Mus
 musculus cDNA clone L0934B10 3', mRNA sequence.

ACCESSION BM120065
 VERSION BM120065.1 GI:17088079
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 546)
 AUTHORS Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Pantano, S., Lim, M.K.
 and Ko, M.S.H.

TITLE Systematic Analyses of NIA Mouse Newborn Kidney cDNA Library
 JOURNAL Unpublished (2001)

COMMENT

Contact: Dawood B. Dudekula
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@igsun.grc.nia.nih.gov
 Plate: L0934 row: B column: 10
 Seq primer: -21M13 Forward
 High quality sequence stop: 546
 POLYA=Yes.

FEATURES
source

Location/Qualifiers
 1..546
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="niaEST:L0934B10-3"
 /db_xref="taxon:10090"
 /clone="L0934B10"
 /clone_lib="NIA Mouse Newborn Kidney cDNA Library (Long)"
 /tissue_type="Newborn Kidney"
 /dev_stage="Newborn"
 /lab_host="DH10B"
 /note="Organ: Kidney; Vector: pSPORT1 (Invitrogen);
 Site 1: SalI; Site 2: NotI; Mouse cDNA project by the
 Laboratory of Genetics, National Institute on Aging (NIA),
 Intramural Research Program, NIH
 (http://igsun.grc.nia.nih.gov/cdna). This is a
 long-transcript enriched cDNA library (Ref. Genome Res.
 11:1553-1558 (2001). [PMID:11544199]). In brief
 double-stranded cDNAs were synthesized with an oligo(dT)
 primer [Invitrogen: 5'-
 pGACTAGTCTAGATCGCGAGCGCCGCCCTTTTTTTTTT-3'] from 26
 microgram of total RNA, treated with T4 DNA polymerase,
 and purified by ethanol-precipitation. The cDNAs were
 ligated to Lone-linker Li-Sal4, purified by
 phenol/chloroform, and separated from free linkers by
 Centricion 100. Then, the cDNAs were amplified by
 long-range high fidelity PCR using Ex Taq polymerase
 (Takara) with a primer Sal4-S. The products were purified
 by phenol/chloroform and Centricion 100. The cDNAs were
 digested with SalI and NotI enzymes, and cloned into
 SalI/NotI site of pSPORT1 plasmid vector. The DH10B E.
 coli host was transformed with ligation mixture by the
 standard chemical method. The average insert size is about
 3.0 kb. The library was constructed by Yulan Piao (NIA)."
 121 a 150 c 128 g 147 t

BASE COUNT
ORIGIN

Query Match 48.5%; Score 25.2; DB 13; Length 546;
 Best Local Similarity 71.7%; Pred. No. 54;
 Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 7 GCATTGCCATATTTGTAGGTCACCTGCGAGTATCTGTGCACA 52
 || ||||| || ||||| ||||| ||||| ||||| ||||| ||
 Db 193 GCCTTGCCTTAATGTGACAGCAGCTCCAGCAGGTACCTGTGCCCA 238

RESULT 13
BM941987

LOCUS BM941987 547 bp mRNA linear EST 29-APR-2002
 DEFINITION UI-M-CG0p-bqv-h-06-0-UI.r1 NIH.BMAP_Ret4_S2 Mus musculus cDNA clone
 UI-M-CG0p-bqv-h-06-0-UI 5', mRNA sequence.

ACCESSION BM941987
 VERSION BM941987.1 GI:19401155
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 547)
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477

FEATURES
source

Location/Qualifiers
1..590
/organism="Mus musculus"
/strain="C57BL/6Ncr"
/db_xref="niaEST:K0535B10-3"
/db_xref="taxon:10090"
/clone="K0535B10"
/clone_lib="NIA Mouse Hematopoietic Stem Cell
(Lin-/C-Kit+/Sca-1+) cDNA Library (Long)"
/tissue_type="Hematopoietic Stem Cell (Lin-/C-Kit+/Sca-1+
)"
/dev_stage="Age approx.10 weeks old"
/lab_host="DH10B"
/note="Vector: pSPORT1 (Invitrogen); Site_1: Sali; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
obtained from Drs. Dennis Taub, Dan Longo (National
Institute on Aging, USA), Jonathan Keller (National Cancer
Institute, USA). Double-stranded cDNAs were synthesized
with an Oligo(dT) primer [Invitrogen:
5'-pCAGTACTTCAGATCGCAGCGCGCCCTTTT-3'] from
4.8 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to Lone-linker LL-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with Sali and NotI enzymes
and cloned into Sali/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 2.7 kb. The library was constructed
by Yulan Piao (NIA)."
BASE COUNT 130 a 162 c 142 g 156 t
ORIGIN

Query Match 48.5%; Score 25.2; DB 13; Length 590;
Best Local Similarity 71.7%; Pred. No. 55;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 7 GCATTGCCATATTGTGAGTGCAGTGCAGTAGGTATCTGTGCACA 52
|| ||||| || ||||| ||||| ||||| ||||| ||||| ||
Db 193 GCCTTGCTTAAATGTGACAGCACTGCCAGAGGTACCTGTGGCCA 238

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 08:58:04 ; Search time 67 Seconds
(without alignments)
238.018 Million cell updates/sec

Title: US-09-875-945-3
Perfect score: 52
Sequence: 1 cctctgcattgcatttt.....gcagtagtgatctgtgcaca 52

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*

2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*

4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*

5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*

6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	21.8	41.9	459	4	US-09-221-017B-308
C 2	21.8	41.9	10803	3	US-09-080-044-1
C 3	21.8	41.9	10803	4	US-09-531-857A-1
C 4	20.6	39.6	1781	4	US-09-040-229B-3
C 5	20.2	38.8	500	4	US-09-115-407-26
C 6	20	38.5	927	4	US-09-134-001C-1597
C 7	20	38.5	2495	4	US-09-668-113A-1
C 8	19.8	38.1	963	4	US-09-134-001C-1259
C 9	19.8	38.1	1751	4	US-09-004-393B-3
C 10	19.8	38.1	3495	1	US-08-446-038B-2
C 11	19.8	38.1	3495	1	US-08-446-010B-2
C 12	19.8	38.1	3495	1	US-08-805-445-2
C 13	19.8	38.1	3495	2	US-08-064-067D-2
C 14	19.8	38.1	3495	2	US-08-066-208-2
C 15	19.8	38.1	3495	4	US-08-980-080-3
C 16	19.8	38.1	3629	1	US-08-097-997A-8
C 17	19.8	38.1	3629	3	US-08-665-574C-8
C 18	19.8	38.1	3629	4	US-08-946-994-8
C 19	19.6	37.7	114	3	US-09-035-648-13
C 20	19.6	37.7	114	4	US-09-001-951-13
C 21	19.6	37.7	114	4	US-08-818-829-13
C 22	19.6	37.7	718	4	US-08-998-416-1133
C 23	19.6	37.7	969	4	US-09-134-001C-2062
C 24	19.6	37.7	3131	3	US-09-035-648-23
C 25	19.6	37.7	3131	4	US-09-001-951-23
C 26	19.6	37.7	3131	4	US-08-818-829-23
C 27	19.6	37.7	6030	1	US-08-441-139-8

C 28	19.4	37.3	179	1	US-08-222-177A-26	Sequence 26, Appl
C 29	19.4	37.3	751	4	US-09-221-017B-18	Sequence 18, Appl
C 30	19.4	37.3	879	5	PCT-US96-05320A-364	Sequence 364, App
C 31	19.4	37.3	1929	4	US-09-337-307A-1	Sequence 1, Appl
C 32	19.4	37.3	2295	3	US-09-189-035-3	Sequence 3, Appl
C 33	19.4	37.3	2295	4	US-09-382-086-3	Sequence 3, Appl
C 34	19.4	37.3	3989	1	US-08-327-494A-1	Sequence 3, Appl
C 35	19.4	37.3	3989	1	US-08-327-494A-3	Sequence 3, Appl
C 36	19.4	37.3	3989	5	PCT-US95-13659-1	Sequence 3, Appl
C 37	19.4	37.3	3989	5	PCT-US95-13659-3	Sequence 3, Appl
C 38	19.4	37.3	4629	2	US-08-484-891-7	Sequence 7, Appl
C 39	19.4	37.3	4670	3	US-08-717-294-41	Sequence 41, Appl
C 40	19.4	37.3	4999	4	US-09-470-618-14	Sequence 14, Appl
C 41	19.4	37.3	4999	4	US-09-364-862-14	Sequence 14, Appl
C 42	19.4	37.3	5035	2	US-08-882-083-1	Sequence 1, Appl
C 43	19.4	37.3	5035	2	US-08-558-107-1	Sequence 1, Appl
C 44	19.4	37.3	5035	3	US-09-243-539-1	Sequence 1, Appl
C 45	19.4	37.3	5793	4	US-09-221-017B-834	Sequence 834, App

ALIGNMENTS

RESULT 1

US-09-221-017B-308/c

; Sequence 308, Application US/09221017B

; Patent No. 6444799

; GENERAL INFORMATION:

; APPLICANT: Ross, Bruce C.

; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

; NUMBER OF SEQUENCES: 1120

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 PAGE MILL ROAD

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304-1018

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows

; SOFTWARE: FastSeq for Windows Version 2.0b

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/221.017B

; FILING DATE: 23-DEC-1998

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PP1182

; FILING DATE: 31-DEC-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PP1546

; FILING DATE: 30-JAN-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PP2911

; FILING DATE: 09-APR-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/AU98/01023

; FILING DATE: 10-DEC-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Monroy, Gladys H

; REGISTRATION NUMBER: 32,430

; REFERENCE/DOCKET NUMBER: 27340-20021.00

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-813-5600

; TELEFAX: 650-494-0792

; TELEX: 706141

; INFORMATION FOR SEQ ID NO: 308:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 459 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: circular

```
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...459
US-09-221-017B-308

Query Match 41.9%; Score 21.8; DB 4; Length 459;
Best Local Similarity 70.7%; Pred. No. 2.7;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 12 GCCATATTGTGAGTCACCTTGCGATGAGTATCTGTGCACA 52
   || ||||| ||| ||||| ||| ||||| ||||| |||
Db 371 GACAGCCTTGCTGTATGCTTGGCGTATGATCTGTGCCCA 331

RESULT 2
US-09-080-044-1/c
; Sequence 1, Application US/09080044
; Patent No. 6074649
; GENERAL INFORMATION:
; APPLICANT: AUDONNET, Jean-Christophe F.
; APPLICANT: BAUDU, Philippe G.
; APPLICANT: RIVIERE, Michel A.
; TITLE OF INVENTION: RECOMBINANT VACCINE CONTAINING FELINE HERPES VIRUS TYPE
; TITLE OF INVENTION: 1, PARTICULARLY FOR TREATING FELINE INFECTIOUS
; TITLE OF INVENTION: PERTONITIS
; FILE REFERENCE: AUDONNET
; CURRENT APPLICATION NUMBER: US/09/080,044
; CURRENT FILING DATE: 1998-05-15
; EARLIER APPLICATION NUMBER: PCT/FR96/01830
; EARLIER FILING DATE: 1996-11-19
; EARLIER APPLICATION NUMBER: 95/14450
; EARLIER FILING DATE: 1995-11-30
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 10803
; TYPE: DNA
; ORGANISM: Feline herpesvirus 1
US-09-080-044-1

Query Match 41.9%; Score 21.8; DB 3; Length 10803;
Best Local Similarity 65.3%; Pred. No. 7.7;
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 CTCCTCGCATTCGCCATATTTGTGAGTGCACCTTGCGATGAGTATCTGTGC 49
   || ||||| ||| ||||| ||| ||||| ||||| |||
Db 748 CACCCCGCATGGCGCTTCTGTGGGAAATTTCCAGCAGCATCATTCG 700

RESULT 3
US-09-531-857A-1/c
; Sequence 1, Application US/09531857A
; Patent No. 6387376
; GENERAL INFORMATION:
; APPLICANT: AUDONNET, Jean-Christophe F.
; APPLICANT: BAUDU, Philippe G.
; APPLICANT: RIVIERE, Michel A.
; TITLE OF INVENTION: RECOMBINANT VACCINE CONTAINING FELINE HERPES VIRUS TYPE
; TITLE OF INVENTION: 1, PARTICULARLY FOR TREATING FELINE INFECTIOUS
; TITLE OF INVENTION: PERTONITIS
; FILE REFERENCE: 454313-2170.1
; CURRENT APPLICATION NUMBER: US/09/531.857A
; CURRENT FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 09/080,044
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: PCT/FR96/01830
; PRIOR FILING DATE: 1996-11-19
; PRIOR APPLICATION NUMBER: 95/14450
US-09-531-857A-1

; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: Feline herpesvirus 1
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...459
US-09-531-857A-1

Query Match 41.9%; Score 21.8; DB 4; Length 10803;
Best Local Similarity 65.3%; Pred. No. 7.7;
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 CTCCTCGCATTCGCCATATTTGTGAGTGCACCTTGCGATGAGTATCTGTGC 49
   || ||||| ||| ||||| ||| ||||| ||||| |||
Db 748 CACCCCGCATGGCGCTTCTGTGGGAAATTTCCAGCAGCATCATTCG 700

RESULT 4
US-09-040-229B-3/c
; Sequence 3, Application US/09040229B
; Patent No. 6432410
; GENERAL INFORMATION:
; APPLICANT: Harland, Richard
; TITLE OF INVENTION: Morphogenic Proteins
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,229B
; FILING DATE: 13-Mar-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B97-020-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1781 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 66..611
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-040-229B-3

Query Match 39.6%; Score 20.6; DB 4; Length 1781;
Best Local Similarity 74.3%; Pred. No. 14;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TCCTCGCATTCGCCATATTTGTGAGTGCACCTTCACGTCGACG 36
   ||||| ||||| ||||| ||||| ||||| |||||
Db 1165 TCCTCAAACTGCCATATTTGCCAGGTTTGTGGCAG 1131

RESULT 5
```

US-09-668-113A-1/c
; Sequence 1, Application US/09668113A

; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0.1

[illegible]

APPLICANT: Wilks, Andrew F.; Ziemiecki, Andrew;
APPLICANT: Harpur, Ailsa
TITLE OF INVENTION: No. 5821069e1 Protein Tyrosine Kinase
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805.445
FILING DATE: 25-FEB-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/446.038
FILING DATE: 19-MAY-1995
APPLICATION NUMBER: 08/064.067
FILING DATE: 30-Jun-1993
APPLICATION NUMBER: PCT/US91/08889
FILING DATE: 26-No. 5821069-1991
APPLICATION NUMBER: Australian PK3594/90
FILING DATE: 28-No. 5821069-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: Australian 88229/91
FILING DATE: 27-No. 5821069-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5821069man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5244
TELEPHONE: 212-688-9200
TELEFAX: 212-838-3884
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3495 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: nucleic acid
US-08-805-445-2

Query Match 38.1%; Score 19.8; DB 1; Length 3495;
Best Local Similarity 63.8%; Pred. No. 39;
Matches 30; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
Qy 2 TCCTGCGATTGCCATATTTGTGAGCTCACTTGCAGTAGGTATCTGTG 48
Db 2486 TCCATGCCCTTGATATCTGATGATGTTTGAAGAAGTTTTTTGTG 2440

RESULT 13
US-08-064-067D-2/c
Sequence 2, Application US/08064067D
Patent No. 5852184
GENERAL INFORMATION:
APPLICANT: Wilks, Andrew F.; Ziemiecki, Andrew;
APPLICANT: Harpur, Ailsa
TITLE OF INVENTION: No. 5852184e1 Protein Tyrosine Kinase
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA

ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/064.067D
FILING DATE: 30-Jun-1993
APPLICATION NUMBER: PCT/US91/08889
FILING DATE: 26-No. 5852184-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: Australian PK3594/90
FILING DATE: 28-No. 5852184-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: Australian 88229/91
FILING DATE: 27-No. 5852184-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5852184man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5244
TELEPHONE: 212-688-9200
TELEFAX: 212-838-3884
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3495 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: nucleic acid
US-08-064-067D-2

Query Match 38.1%; Score 19.8; DB 2; Length 3495;
Best Local Similarity 63.8%; Pred. No. 39;
Matches 30; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
Qy 2 TCCTGCGATTGCCATATTTGTGAGCTCACTTGCAGTAGGTATCTGTG 48
Db 2486 TCCATGCCCTTGATATCTGATGATGTTTGAAGAAGTTTTTTGTG 2440

RESULT 14
US-09-066-208-2/c
Sequence 2, Application US/09066208
Patent No. 5910426
GENERAL INFORMATION:
APPLICANT: Wilks, Andrew F.; Ziemiecki, Andrew;
APPLICANT: Harpur, Ailsa
TITLE OF INVENTION: No. 5910426e1 Protein Tyrosine Kinase
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066.208
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/805.445
FILING DATE: 25-FEB-1997
APPLICATION NUMBER: US 08/446.038
FILING DATE: 19-MAY-1995

```
; APPLICATION NUMBER: 08/064,067
; FILING DATE: 30-Jun-1993
; APPLICATION NUMBER: PCT/US91/08889
; FILING DATE: 26-No. 5910426-1991
; PRIOR APPLICATION DATA: Australian PK3594/90
; APPLICATION NUMBER: Australian PK3594/90
; FILING DATE: 28-No. 5910426-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: Australian 88229/91
; FILING DATE: 27-No. 5910426-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5910426man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5244
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-688-9200
; TELEFAX: 212-838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3495 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: nucleic acid
; US-09-066-208-2
```

```
Query Match 38.1%; Score 19.8; DB 2; Length 3495;
Best Local Similarity 63.8%; Pred. No. 39;
Matches 30; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
```

```
QY 2 TCCTCGCATGCCATATTTGTGAGGTCACCTGCGAGTAGGATCTGTG 48
   ||| || | |||| || || || || || || || || || || || || ||
Db 2486 TCCATGCCCTTGCATATCTGAGATGTGTATTGAAGAAGTTTTTTGTG 2440
```

```
RESULT 15
US-08-980-080-3/c
; Sequence 3, Application US/08980080
; Patent No. 6312941
; GENERAL INFORMATION:
; APPLICANT: CARTER-SU, CHRISTIN
; APPLICANT: RUI, LIANG-YOU
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
; TITLE OF INVENTION: SIGNALING PATHWAY AGONISTS AND ANTAGONISTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/980,080
; FILING DATE: 26-NOV-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: UM-03036
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3495 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2979
; US-08-980-080-3
Query Match 38.1%; Score 19.8; DB 4; Length 3495;
Best Local Similarity 63.8%; Pred. No. 39;
Matches 30; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 2 TCCTCGCATGCCATATTTGTGAGGTCACCTGCGAGTAGGATCTGTG 48
   ||| || | |||| || || || || || || || || || || || || ||
Db 2486 TCCATGCCCTTGCATATCTGAGATGTGTATTGAAGAAGTTTTTTGTG 2440
Search completed: November 26, 2002, 10:13:00
Job time : 73 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 10:13:04 ; Search time 32 Seconds
(without alignments)
625.813 Million cell updates/sec

Title: US-09-875-945-3
Perfect score: 52
Sequence: 1 ctctcgcattgccatattt.....gcagtagtatctgtgcaca 52

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 341543 seqs, 192557720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_applications_NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	52	100.0	52	10	US-09-875-945-3
2	23.4	45.0	560	10	US-09-864-761-7874
3	23.2	44.6	1647	9	US-09-938-842A-19
4	22.4	43.1	217	10	US-09-983-965-5516
5	22.4	43.1	415	10	US-09-983-965-5516
6	21.8	41.9	482	10	US-09-764-847-455
7	21.8	41.9	26048	10	US-09-764-869-1556
8	21.4	41.2	1343	10	US-09-887-576-505
9	21.4	41.2	1343	10	US-09-887-576-505
10	21.4	41.2	2000	9	US-09-938-842A-3988
11	21.4	41.2	2003	10	US-09-887-576-189
12	21.4	41.2	202001	10	US-09-734-674-3
13	21.2	40.8	180216	10	US-09-835-232-6
14	21	40.4	973	10	US-09-925-297-195
15	20.8	40.0	1646	10	US-09-925-300-408
16	20.8	40.0	1795	10	US-09-925-300-19
17	20.8	40.0	6074	10	US-09-880-107-2378
18	20.6	39.6	474	10	US-09-770-444-84
19	20.6	39.6	1387	10	US-09-764-846-113

20	20.6	39.6	31168	9	US-09-764-868-1464	Sequence 1464, Ap
21	20.4	39.2	276	10	US-09-878-574-8147	Sequence 8147, Ap
22	20.4	39.2	2299	10	US-09-925-300-564	Sequence 564, App
23	20.4	39.2	2528	10	US-09-925-300-582	Sequence 582, App
24	20.4	39.2	4043	10	US-09-954-456-1918	Sequence 1918, Ap
25	20.2	38.8	426	10	US-09-924-035A-598	Sequence 598, App
26	20.2	38.8	4059	12	US-10-042-417-47	Sequence 47, Appl
27	20.2	38.8	10126	10	US-09-764-877-2194	Sequence 2194, Ap
28	20	38.5	396	10	US-09-878-574-889	Sequence 889, App
29	20	38.5	455	10	US-09-998-598-438	Sequence 438, App
30	20	38.5	557	10	US-09-864-761-13310	Sequence 13310, A
31	20	38.5	572	10	US-09-777-564-83	Sequence 83, Appl
32	20	38.5	738	10	US-09-919-580-799	Sequence 799, App
33	20	38.5	1302	10	US-09-772-105-56	Sequence 56, Appl
34	20	38.5	2808	10	US-09-905-983-59	Sequence 59, Appl
35	19.8	38.1	29220	9	US-09-764-868-1312	Sequence 1312, Ap
36	19.8	38.1	29220	9	US-09-764-868-1313	Sequence 1313, Ap
37	19.6	37.7	229	10	US-09-728-446-113	Sequence 113, App
38	19.6	37.7	327	10	US-09-783-590-1998	Sequence 1998, Ap
39	19.6	37.7	400	10	US-09-789-919-35	Sequence 35, Appl
40	19.6	37.7	415	10	US-09-983-965-629	Sequence 629, App
41	19.6	37.7	486	10	US-09-864-761-15043	Sequence 15043, A
42	19.6	37.7	910	12	US-10-044-090-728	Sequence 728, App
43	19.6	37.7	1026	10	US-09-815-242-7701	Sequence 7701, Ap
44	19.6	37.7	2000	9	US-09-938-842A-3534	Sequence 3534, Ap
45	19.6	37.7	2623	9	US-09-804-409A-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-875-945-3
; Sequence 3, Application US/09875945
; Patent No. US20020098169A1
; GENERAL INFORMATION:
; APPLICANT: METCON MEDICIN AB
; APPLICANT: SMITH, Ulf
; TITLE OF INVENTION: No. US20020098169A1el sequences and their use
; FILE REFERENCE: 45513MH
; CURRENT APPLICATION NUMBER: US/09/875,945
; PRIOR FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: SE 0002189-9
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: US 60/210,207
; PRIOR FILING DATE: 2000-06-08
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; TYPE: DNA
; LENGTH: 52
; ORGANISM: Homo sapiens
US-09-875-945-3

Query Match	100.0%;	Score 52;	DB 10;	Length 52;
Best Local Similarity	100.0%;	Pred. No. 1.6e-12;		
Matches	52;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;

QY 1 CTCCTCGCATTCGCATATTGTGAGTCACTTGCGAGTAGGTACTGTGTCACA 52
|||||
Db 1 CTCCTCGCATTCGCATATTGTGAGTCACTTGCGAGTAGGTACTGTGTCACA 52

RESULT 2
US-09-864-761-7874
; Sequence 7874, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO

2
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842

GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

```

; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 5116
; LENGTH: 415
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 33-LIB34-008-Q1-E1-A2
US-09-983-965-5116

Query Match      43.1%; Score 22.4; DB 10; Length 415;
Best Local Similarity 81.2%; Pred. No. 4.3;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy  2  TCCTCGCATTCGCATATTTGTGAGGTCACCTTG 33
    ||| |||| | ||||| ||||| ||| || |
Db  306 TCCCGCGCACTACCATATTTGTGAAGTCGCTGG 337

RESULT 6
US-09-764-847-455
; Sequence 455, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 455
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (468)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-847-455

Query Match      41.9%; Score 21.8; DB 10; Length 482;
Best Local Similarity 70.7%; Pred. No. 8;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy  10  TTGCCATATTTGTGAGGTCACCTTCGACGTAGTATCTGTGCA 50
    ||||| ||||| ||||| ||| ||| ||| |||||
Db  319 TTGCCATTTGTGTAAGTCCTGTAGAGAACTTCTCTGTGCA 359

RESULT 7
US-09-764-869-1556/c
; Sequence 1556, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper

```

```

; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1556
; LENGTH: 26048
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-869-1556

Query Match      41.9%; Score 21.8; DB 10; Length 26048;
Best Local Similarity 78.8%; Pred. No. 28;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy  1  CTCCTCGCATTCGCATATTTGTGAGGTCACCTTG 33
    ||| ||| | ||||| ||||| ||||| ||| ||
Db  13166 CTCCTCCCACTACCATATATGTGAGGACACGCTG 13134

RESULT 8
US-09-887-576-505
; Sequence 505, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 505
; LENGTH: 1343
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-887-576-505

Query Match      41.2%; Score 21.4; DB 10; Length 1343;
Best Local Similarity 71.8%; Pred. No. 16;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy  9  ATTGCCATATTTGTGAGGTCACCTTCGACGTAGTATCTGT 47
    ||||| ||||| ||||| ||||| ||||| |||||
Db  688 ATTAGATTTTGTGAATTCATTACAGTAGAAAACCTGT 726

RESULT 9
US-09-887-576-563
; Sequence 563, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848

```

; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 563
; LENGTH: 1343
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-887-576-563

Query Match 41.2%; Score 21.4; DB 10; Length 1343;
Best Local Similarity 71.8%; Pred. No. 16;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 9 ATTGCCATATTTGTGAGTGCACCTTGCAGTAGGTATCTGT 47
||| ||||| ||||| ||||| ||||| ||||| |||||
Db 688 ATTTAGATTTTGTGAATTCATTACAGTAGAAAACTGT 726

RESULT 10

US-09-938-842A-3988
; Sequence 3988, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPL300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3988
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3988

Query Match 41.2%; Score 21.4; DB 9; Length 2000;
Best Local Similarity 71.8%; Pred. No. 18;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 9 ATTGCCATATTTGTGAGTGCACCTTGCAGTAGGTATCTGT 47
||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1344 ATTTAGATTTTGTGAATTCATTACAGTAGAAAACTGT 1382

RESULT 11

US-09-887-576-189
; Sequence 189, Application US/09887576
; Patent No. US2002014047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576

; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 189
; LENGTH: 2003
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-887-576-189

Query Match 41.2%; Score 21.4; DB 10; Length 2003;
Best Local Similarity 71.8%; Pred. No. 18;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 9 ATTGCCATATTTGTGAGTGCACCTTGCAGTAGGTATCTGT 47
||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1343 ATTTAGATTTTGTGAATTCATTACAGTAGAAAACTGT 1381

RESULT 12

US-09-734-674-3
; Sequence 3, Application US/09734674
; Patent No. US20020081648A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001018
; CURRENT APPLICATION NUMBER: US/09/734,674
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 202001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(202001)
; OTHER INFORMATION: n = A,T,C or G
US-09-734-674-3

Query Match 41.2%; Score 21.4; DB 10; Length 202001;
Best Local Similarity 71.8%; Pred. No. 76;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 12 GCCATATTTGTGAGTGCACCTTGCAGTAGGTATCTGTGCA 50
||| ||||| ||||| ||||| ||||| ||||| |||||
Db 49859 GCTAGATATATGTTGTTAATTCAGTATGAATCTGTGAA 49897

RESULT 13

US-09-835-232-6/c
; Sequence 6, Application US/09835232
; Patent No. US20020098489A1
; GENERAL INFORMATION:
; APPLICANT: Leder, Philip
; APPLICANT: Leader, Benjamin
; TITLE OF INVENTION: FORMIN-2 NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 00383/052002
; CURRENT APPLICATION NUMBER: US/09/835,232
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,811
; PRIOR FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0

```

; SEQ ID NO 6
; LENGTH: 180216
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(180216)
; OTHER INFORMATION: n = A,T,C or G
US-09-835-232-6

Query Match          40.8%; Score 21.2; DB 10; Length 180216;
Best Local Similarity 76.5%; Pred. No. 89;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 TCCTCGCATTTGAGGTCACTTGCA 35
   || ||||| |||| ||||| |||||
Db 21479 TCATGTCATTCCTTATTCATGAGTCCCTTTCA 21446

RESULT 14
US-09-925-297-195
; Sequence 195, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 195
; LENGTH: 973
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (88)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (89)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (101)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (189)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (895)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (960)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (969)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-297-195

Query Match          40.4%; Score 21; DB 10; Length 973;
Best Local Similarity 66.7%; Pred. No. 21;
Matches 30; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 8 CATTGCCATATTTGAGGTCACTTGCACTGTCGACA 52
   ||||| ||||| ||||| ||||| ||||| |||||
Db 419 CATTGCCATATTTGAGGTCACTTGCACTGTCGACA 463

RESULT 15
```

```

US-09-925-300-408/c
; Sequence 408, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 408
; LENGTH: 1646
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (55)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-408

Query Match          40.0%; Score 20.8; DB 10; Length 1646;
Best Local Similarity 64.6%; Pred. No. 30;
Matches 31; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 5 TCGCATTGCCATATTTGAGGTCACTTGCACTGTCGACA 52
   ||||| ||||| ||||| ||||| ||||| |||||
Db 1092 TTGAATTGCTTTATTTGTGGGTTTCTTCTAAGAGTTAATTGGAACA 1045

Search completed: November 26, 2002, 14:33:07
Job time : 108 secs
```

THIS PAGE BLANK (USPTO)